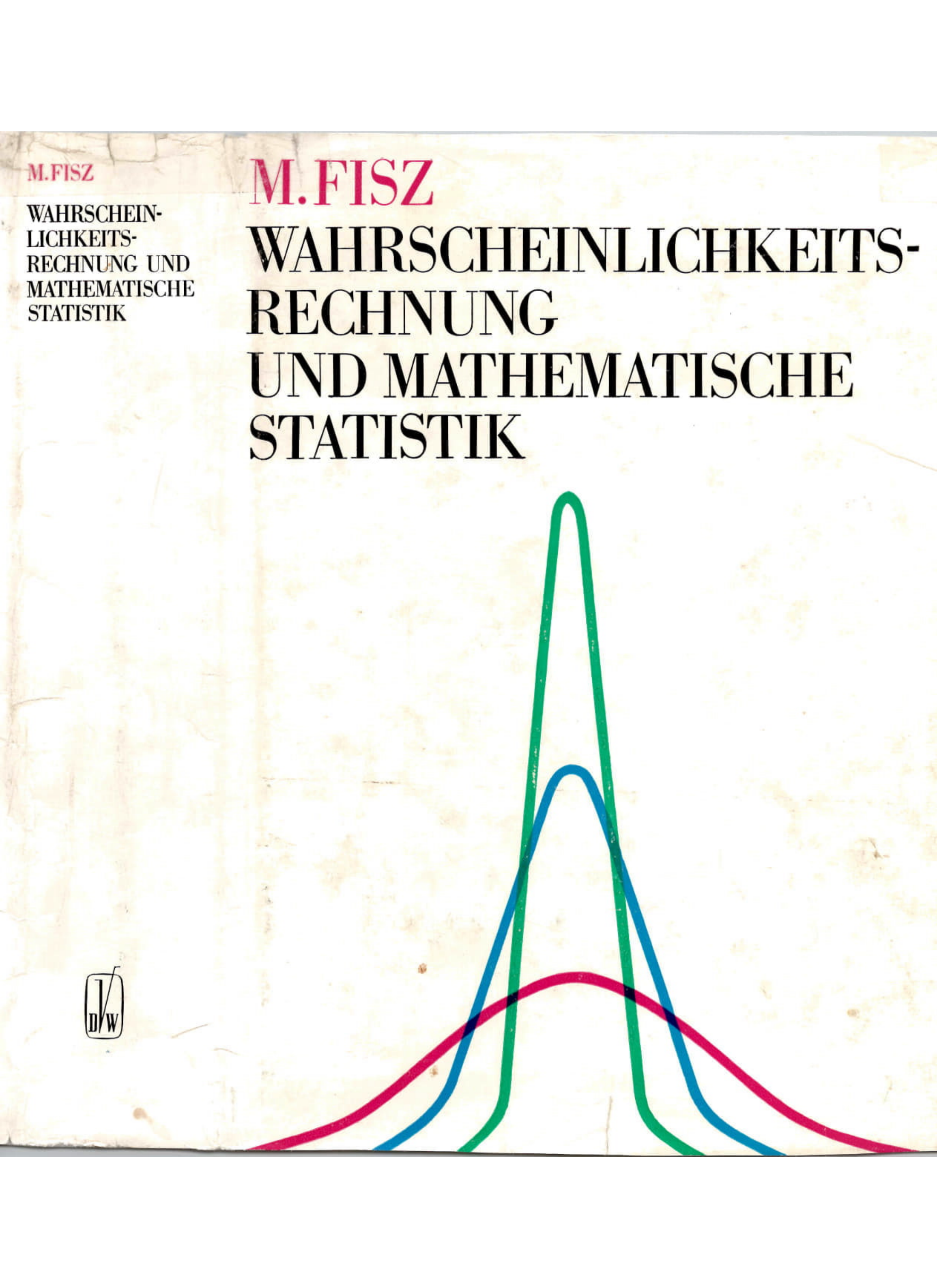
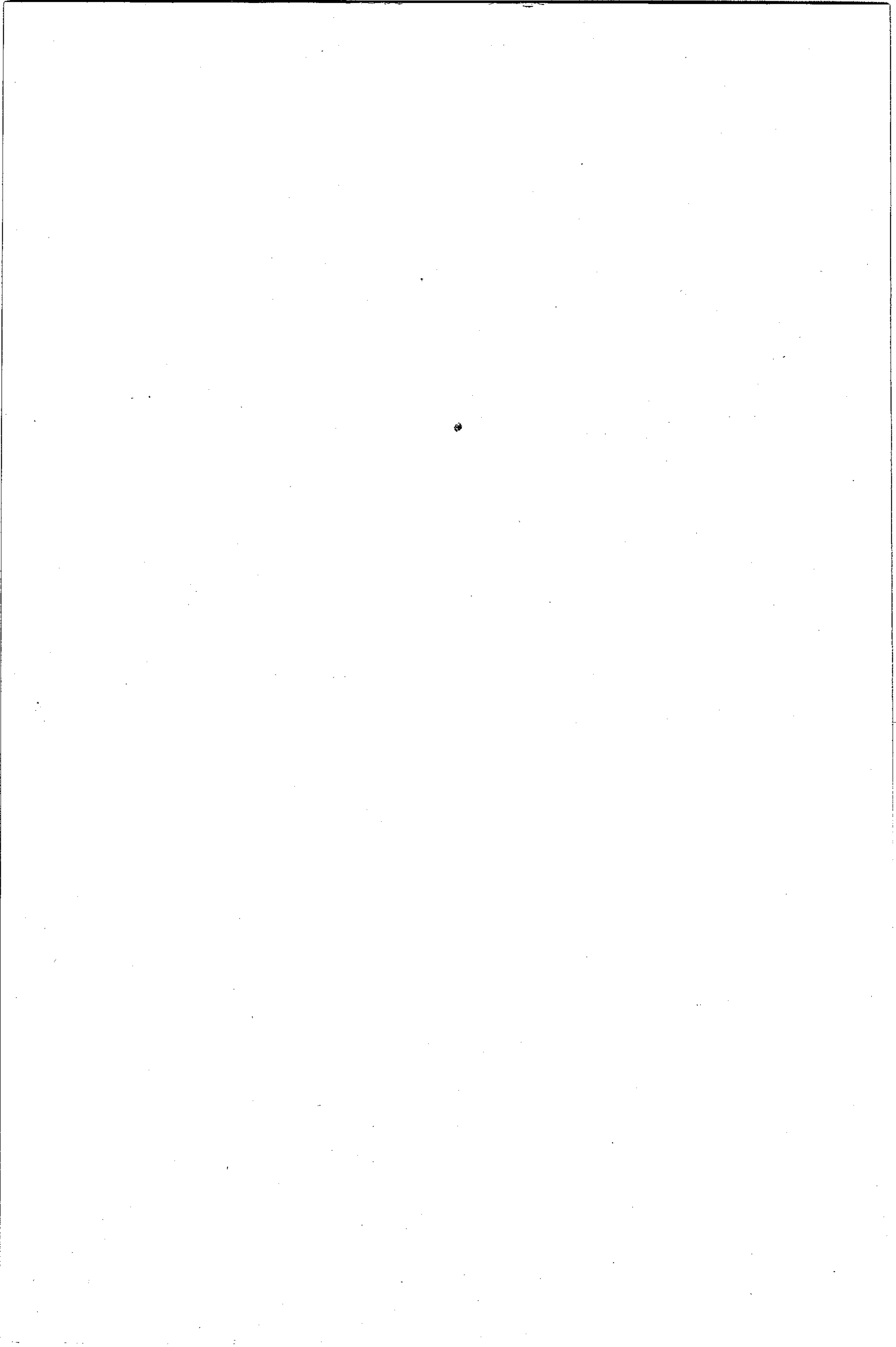

M. FISZ
WAHRSCHEIN-
LICHKEITS-
RECHNUNG UND
MATHEMATISCHE
STATISTIK
M. FISZ
WAHRSCHEINLICHKEITS-
RECHNUNG
UND MATHEMATISCHE
STATISTIK



M. Fisz · Wahrscheinlichkeitsrechnung und mathematische Statistik

Hochschulbücher für Mathematik

Band 40

Herausgegeben von H. Grell, K. Maruhn und W. Rinow

Wahrscheinlichkeitsrechnung und mathematische Statistik

von Marek Fisz

Mit 37 Abbildungen, 40 Tabellen und 8 Tafeln
Achte Auflage



VEB Deutscher Verlag der Wissenschaften
Berlin 1976

Marek Fisz
Rachunek prawdopodobieństwa i statystyka matematyczna
Wydanie trzecie, poprawione i rozszerzone
Państwowe Wydawnictwo Naukowe, Warszawa 1967

Autorisierte Übersetzung aus dem Polnischen
und wissenschaftliche Redaktion:
Hannelore und Rolf Sulanke, Josef Wloka, Viktor Ziegler
Verantwortliche Verlagsredakteure: Erika Arndt, Brigitte Mai

Printed in the German Democratic Republic
Lizenz-Nr. 206 · 435/180/76
Einband und Schutzumschlag: Rudolf Wendt, Berlin
Satz: VEB Druckhaus „Maxim Gorki“, Altenburg
Offsetnachdruck und buchbinderische Verarbeitung:
VEB Druckerei „Thomas Müntzer“, 582 Bad Langensalza
LSV 1074
Bestellnummer: 569 759 3
EVP 45,— Mark

VORWORT

Das vorliegende Buch beginnt mit dem Satz: „Die Wahrscheinlichkeitsrechnung ist ein Teilgebiet der Mathematik, das sich damit befaßt, die Gesetzmäßigkeiten unter den zufälligen Ereignissen aufzudecken und zu untersuchen.“ Doch wäre man noch vor einigen Jahrzehnten kaum geneigt gewesen, diesen Satz für akzeptabel (oder gar für evident) zu halten, weder seitens der Mathematiker noch seitens der Forscher, die sich der Wahrscheinlichkeitsrechnung bedient haben. Erst in den zwanziger und dreißiger Jahren unseres Jahrhunderts wurden das Wesen der Wahrscheinlichkeitsrechnung als eines Zweiges der Mathematik erkannt und die Beziehungen zwischen den Begriffen der „Wahrscheinlichkeit“ und der „Häufigkeit“ zufälliger Ereignisse geklärt. In 1.3. wird der Leser in einer umfangreichen (aber sicher unvollständigen) Liste Namen von Forschern finden, deren Beiträge für das Gebiet der Grundlagen der Wahrscheinlichkeitsrechnung wichtig sind. Jedoch sollte die besondere Bedeutung der „Théorie Analytique des Probabilités“ von LAPLACE, der „Wahrscheinlichkeit, Statistik und Wahrheit“ von v. MISES und der „Grundbegriffe der Wahrscheinlichkeitsrechnung“ von KOLMOGOROFF besonders hervorgehoben werden. Jedes dieser drei Werke leitete jeweils einen neuen Abschnitt in der Geschichte der Wahrscheinlichkeitsrechnung ein. Darüber hinaus haben die Arbeiten der Steinhausschen Schule der unabhängigen Funktionen sehr viel zur Klärung der Grundbegriffe der Wahrscheinlichkeitsrechnung beigetragen.

Der Fortschritt auf dem Gebiet der Grundlagen der Wahrscheinlichkeitsrechnung sowie die Einführung der Theorie der charakteristischen Funktionen begünstigten die ungewöhnlich rasche Entwicklung der modernen Wahrscheinlichkeitsrechnung. Im Bereich der Grenzwertsätze für Summen von unabhängigen Zufallsvariablen wurde eine sehr allgemeine Theorie entwickelt (CHINTSCHIN, LÉVY, KOLMOGOROFF, FELLER, GNEDENKO u. a.), während für abhängige Zufallsvariable einige wichtige Teilergebnisse gewonnen wurden (BERNSTEIN, MARKOFF, DOEBLIN, KOLMOGOROFF u. a.). Darüber hinaus wurde die Theorie der stochastischen Prozesse zu einem exakten mathematischen Zweig der Wahrscheinlichkeitsrechnung (MARKOFF, CHINTSCHIN, LÉVY, WIENER, KOLMOGOROFF, FELLER, CRAMÉR, DOOB u. a.).

Manche Ideen der Anwendung der Wahrscheinlichkeitsrechnung, die heute in der mathematischen Statistik ihren Niederschlag finden, gehen auf BAYES (Theorie der Schätzungen), LAPLACE (Qualitätskontrolle von Arzneimitteln) und

Gauss (Fehlerrechnung) zurück. Doch erst in unserem Jahrhundert wurde die mathematische Statistik zu einem eigenständigen Gebiet der Wissenschaft. Um nur einige der wichtigsten Persönlichkeiten zu nennen, denen diese Entwicklung zu verdanken ist, erwähnen wir K. Pearson, R. A. Fisher, J. Neyman und A. Wald, deren Ideen und systematische Forschungen sehr viel zu dem hohen Entwicklungsstand der modernen mathematischen Statistik beigetragen haben.

Gegenwärtig schreitet die Entwicklung der Wahrscheinlichkeitsrechnung und der mathematischen Statistik sehr intensiv voran. Einerseits ziehen bisher ungelöste Probleme der klassischen Wahrscheinlichkeitsrechnung die Aufmerksamkeit auf sich, andererseits werden große Anstrengungen unternommen, um sehr weitgehende Verallgemeinerungen alter Begriffe zu gewinnen, insbesondere durch Betrachtung der Wahrscheinlichkeitsrechnung in allgemeineren als den üblicherweise betrachteten endlichdimensionalen euklidischen Räumen. Wie eng die Wahrscheinlichkeitsrechnung gegenwärtig mit anderen Teilgebieten der Mathematik verknüpft ist, geht vor allen Dingen daraus hervor, daß fast unmittelbar nach der Formulierung der Theorie der Distributionen durch L. Schwartz und J. Mikusiński die zugehörigen wahrscheinlichkeitstheoretischen Gegenstücke gründlich erörtert wurden (Gelfand, Itô, Urbanik). Darüber hinaus sind Wahrscheinlichkeitsrechnung und mathematische Statistik nicht mehr nur einfach „Kunden" anderer Teile der Mathematik. Im Gegenteil, die gegenseitige Beeinflussung und Durchdringung von Wahrscheinlichkeitsrechnung und mathematischer Statistik und anderen Gebieten der Mathematik schreitet rasch voran. Ein Beispiel dafür ist die Beziehung zwischen der analytischen Zahlentheorie und der Wahrscheinlichkeitsrechnung (Borel, Chintschin, Linnik, Erdös, Kac, Rényi u. a.), ein anderes die Anwendung der Spieltheorie in der mathematischen Statistik und umgekehrt deren stimulierender Einfluß auf die Spieltheorie (v. Neumann, Morgenstern, Wald, Blackwell, Karlin u. a.).

Die Anwendungsgebiete der Wahrscheinlichkeitsrechnung und der mathematischen Statistik erweitern sich ständig. Statistische Verfahren werden heute vielfach in Physik, Biologie, Medizin, Ökonomie, Industrie, Landwirtschaft, Fischereiwesen, Meteorologie und im Verkehrswesen angewandt. Die statistische Methodologie wurde zu einer wichtigen Komponente der wissenschaftlichen Schlußweise und zu einem untrennbaren Bestandteil einer gut organisierten Arbeit in Betrieben und Institutionen.

Die Entwicklung von Wahrscheinlichkeitsrechnung und mathematischer Statistik sowie ihrer Anwendungen äußert sich in einer ständig wachsenden Flut wissenschaftlicher Veröffentlichungen.

Nach diesem kurzen Überblick über die Stellung der modernen Wahrscheinlichkeitsrechnung und mathematischen Statistik möchte ich nun auf das Hauptanliegen des vorliegenden Buches eingehen; es stellt sich folgende Aufgaben:

1. Darbietung einer systematischen Einführung in die moderne Wahrscheinlichkeitsrechnung und mathematische Statistik.

2. Vermittlung einer Vorstellung von der Vielfalt der möglichen Anwendungen dieser Theorien, untermalt durch anschauliche konkrete Beispiele.
3. Ausführliche (wenn auch nicht erschöpfende) Bezugnahme auf andere Bücher und Arbeiten, meist mit kurzen Angaben über deren Inhalt, um so dem Leser die Vervollständigung seines Wissens über die betrachteten Gegenstände zu erleichtern.

Trotz aller Sorgfalt, die darauf verwandt wurde, das Buch mathematisch streng aufzubauen, wurde großer Nachdruck auf eine anschauliche Betrachtungsweise und auf die Anwendbarkeit der Begriffe und Sätze gelegt.

In den meisten Fällen werden die Sätze mit vollständigen Beweisen gebracht. Nur Beweise, die entweder zu langwierig sind oder mathematische Kenntnisse voraussetzen, die weit über die sonst in diesem Buch benötigten hinausgehen, wurden weggelassen.

Der gesamte Text des Buches kann von Studenten gelesen werden, die über gewisse Grundlagen der Analysis und Algebra verfügen. Weitreichende Kenntnisse in diesen Gebieten oder Kenntnisse aus der Maß- und Integrationstheorie werden jedoch nicht benötigt. Einige in diesen Bereich fallende Begriffe (z. B. der des Stieltjes-Integrals) werden im Text erläutert. Darüber hinaus ist dem Buch ein Anhang beigegeben, in dem gewisse Grundbegriffe und Sätze der modernen Maß- und Integrationstheorie behandelt werden.

An jedes Kapitel schließt sich ein Abschnitt „Aufgaben und Ergänzungen" an. Ein großer Teil dieser Aufgaben ist verhältnismäßig einfach und sollte vom Leser gelöst werden. Die übrigen hingegen dienen der Information und Anregung.

Das Buch eignet sich zur Verwendung bei systematischen Einjahreslehrgängen der Wahrscheinlichkeitsrechnung und mathematischen Statistik. Einen Teil des hier gebotenen Stoffes habe ich in meinen jährlich wiederkehrenden Vorlesungszyklen an der Universität Warschau in den Studienjahren 1951/52 bis 1959/60, im Frühjahr 1957 an der Universität Peking und in den letzten Jahren in den Vereinigten Staaten von Amerika an den Universitäten Washington, Stanford, Columbia und New York vorgetragen.

Das Buch ist auch für Nichtmathematiker geeignet, die sich für Begriffe, Sätze und Anwendungsverfahren interessieren.

Mit der Niederschrift des Manuskripts habe ich 1950 begonnen. Seine erste Auflage (374 Seiten) erschien polnisch im Jahre 1954 und war in wenigen Monaten vergriffen. Ich habe sodann die zweite, durchgesehene und erweiterte polnische Auflage vorbereitet, die im Jahre 1958 gleichzeitig mit ihrer deutschen Übersetzung erschien. Im Vergleich zur zweiten Auflage enthält diese dritte Auflage[1]) sehr viele Ergänzungen und Änderungen. Hier die wichtigsten:

Völlig neu sind die Abschnitte „Aufgaben und Ergänzungen", der „Anhang" sowie die Paragraphen bzw. Abschnitte 2.7.C, 2.8.C, 3.2.C, 3.6.G, 4.6.B, 6.4.B,

[1]) Dieses Vorwort schrieb der Verfasser zu der Auflage, die in englischer Sprache 1963 in New York erschien (Anm. d. Red.).

6.4.C, 6.12.D, 6.12.F, 6.15, 8.11, 9.4.B, 9.6.E, 9.9.B, 10.10.B, 10.11.E, 12.6.B, 13.5.E, 13.7.D, 14.2.E, 14.4.D, 15.1.C, 15.3.C, 16.3.D, 16.6 und 17.10.A. Der Abschnitt 8.10 (stationäre Prozesse) wurde ebenfalls fast völlig neu geschrieben.

Erheblich verändert oder ergänzt wurden die Abschnitte 2.5.C, 3.5, 3.6.C, 4.1, 4.2, 5.6.B, 5.7, 5.13.B, 6.2, 6.4.A, 6.5, 6.12.E, 6.12.G, 7.5.B, 8.4.D, 8.8.B, 8.12, 9.1, 9.7, 10.12, 10.13, 12.4.C, 12.4.D, 13.3, 16.2.C.

Diese Änderungen und Ergänzungen wurden vorgenommen, um dem obengenannten Hauptanliegen des Werkes besser gerecht zu werden.

J. Lukaszewicz, A. M. Rusiecki und W. Sadowski haben das Manuskript der ersten Auflage gelesen und viele Verbesserungen angeregt. Die kritischen Bemerkungen von E. Marczewski sowie die von Z. W. Birnbaum und S. Zubrzycki verfaßten Rezensionen der ersten Auflage haben mir gute Dienste bei der Vorbereitung der zweiten Auflage geleistet. Desgleichen waren die wertvollen Bemerkungen von K. Urbanik sehr nützlich, der das Manuskript zur zweiten Auflage gelesen hatte. Zahlreiche Bemerkungen und Berichtigungen wurden angeregt von J. Wojtyniak und R. Zasępa (erste Auflage) und von L. Kubik, R. Sulanke und J. Włoka (zweite Auflage). Ihnen allen bin ich in tiefstem Dank verbunden.

New York, Oktober 1962 Marek Fisz

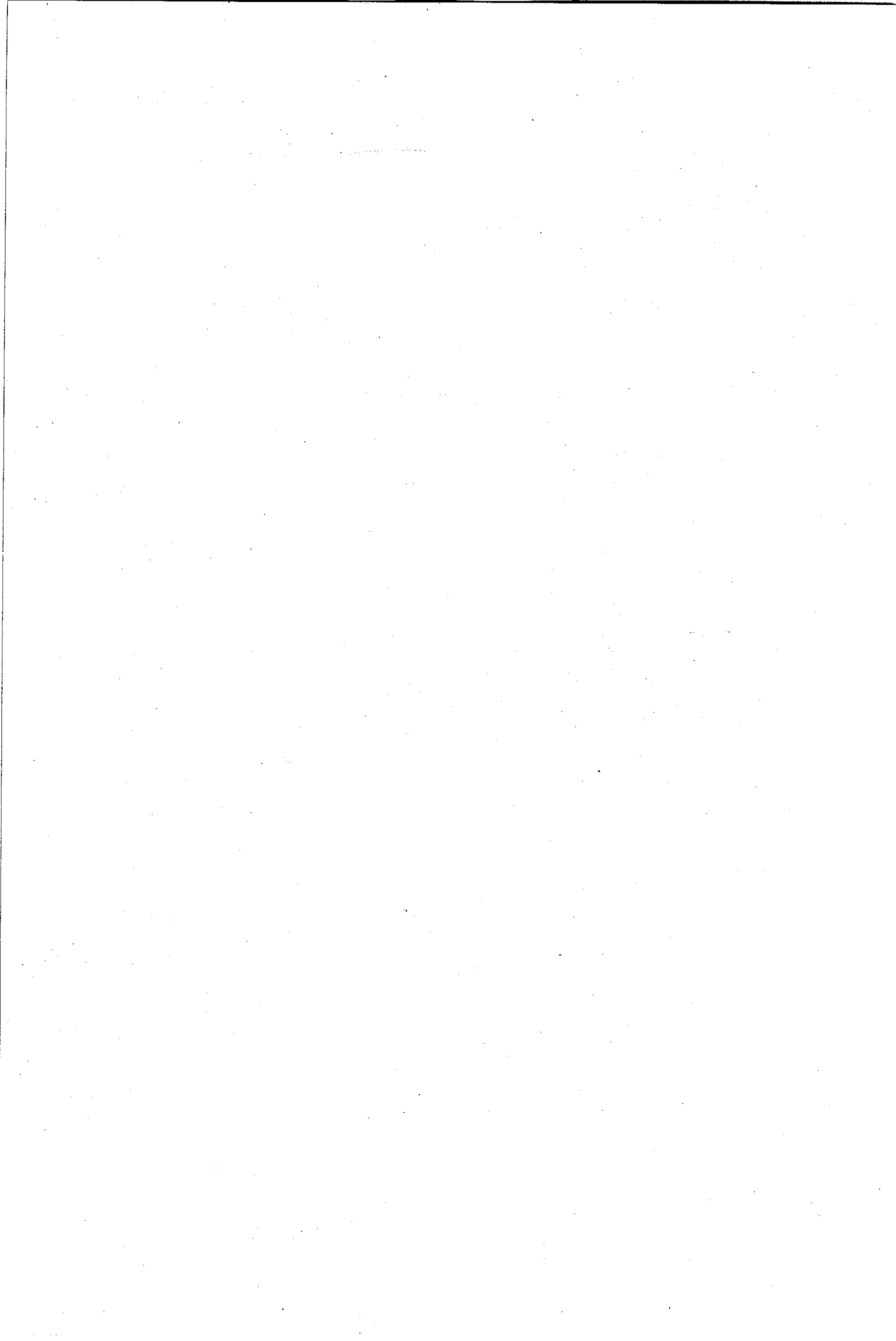
INHALT

Wahrscheinlichkeitsrechnung

1. Zufällige Ereignisse

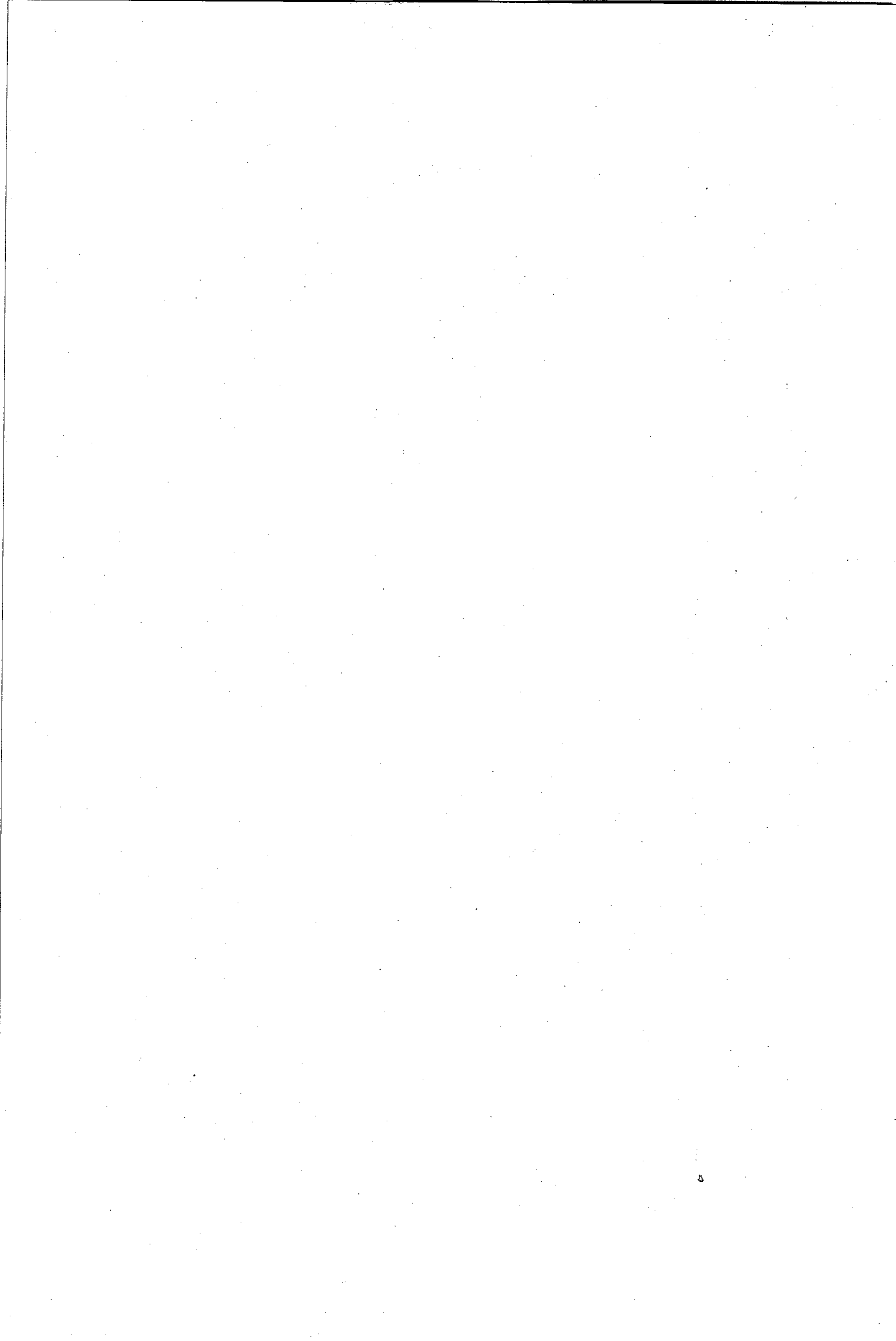
2. Zufallsvariable

3. Die Parameter der Verteilung einer Zufallsvariablen



ERSTER TEIL

WAHRSCHEINLICHKEITSRECHNUNG



1. ZUFÄLLIGE EREIGNISSE

1.1. Einleitende Bemerkungen

A. Die Wahrscheinlichkeitsrechnung ist ein Teilgebiet der Mathematik, das sich damit befaßt, die Gesetzmäßigkeiten unter den zufälligen Ereignissen aufzudecken und zu untersuchen. An Beispielen soll gezeigt werden, was man unter einem *zufälligen Ereignis* versteht.

Beispiel 1.1.1. Wir werfen ein symmetrisches Geldstück auf den Boden. Es wird entweder die „Zahl" oder der „Adler" oben liegen. Das Ergebnis eines einzelnen Wurfes kann man nicht vorhersagen, obwohl jedes Resultat durch einen gewissen Ursachenkomplex bestimmt wird. Zu diesen Ursachen gehören die Anfangsgeschwindigkeit des Geldstücks, der Wurfwinkel, der Glättegrad der Bodenoberfläche, auf die das Geldstück fällt, usw. Da wir aber diese Ursachen weder beliebig genau regulieren können noch genau kennen, sind wir auch nicht in der Lage, das Ergebnis jedes einzelnen Wurfes im voraus zu berechnen.

Das Ergebnis eines Wurfes mit der Geldmünze — Zahl oder Adler (in der Literatur auch: „Kopf oder Wappen") — bezeichnet man als ein *zufälliges Ereignis*.

Beispiel 1.1.2. Nehmen wir an, daß Jahr für Jahr in einem bestimmten Monat und an einem bestimmten Ort die mittlere Temperatur beobachtet wird, etwa im Januar in Warschau.[1]) Dieser Mittelwert hängt von vielen Ursachen ab, z. B. von der Richtung und Stärke des Windes und vom Feuchtigkeitsgehalt der Luft. Die Auswirkung dieser Ursachen ist von Jahr zu Jahr verschieden; deshalb ist die mittlere Temperatur im Januar in Warschau nicht Jahr für Jahr die gleiche. Zwar können hier die Ursachen für die verschiedenen mittleren Temperaturen festgestellt werden. Häufig ist es aber nicht möglich, die Gründe für das Auftreten dieser Ursachen zu bestimmen, so daß wir im Endergebnis die Temperatur nicht exakt voraussagen können. Wir sprechen deshalb von einem zufälligen Ereignis.

B. Es könnte der Eindruck entstehen, daß sich bei den angegebenen Beispielen keinerlei Gesetzmäßigkeit beobachten läßt. Das ist aber doch möglich, wenn die Anzahl der Beobachtungen groß genug ist, wenn es sich um eine „Massenerscheinung" handelt.

Kehren wir zum Beispiel 1.1.1 zurück. Man kann zwar nicht das Ergebnis eines einzelnen Wurfes voraussagen; wird aber eine lange Serie von Würfen ausgeführt, so bemerkt man, daß der Adler annähernd ebenso häufig wie die Zahl vorkommt. Mit n sei die Anzahl aller Würfe bezeichnet, mit m die Anzahl der Würfe, bei denen der Adler vorkommt. Den Bruch $\frac{m}{n}$ nennen wir die *Häufigkeit* des Adlers. Die Häufigkeit der Zahl ist gleich dem Bruch $\frac{n-m}{n}$. Die Er-

[1]) Siehe Beispiel 12.5.1.

fahrung zeigt, daß bei genügend großen Werten von n, das heißt bei vielen Würfen mit dem Geldstück (also wieder bei einer Massenerscheinung), die Häufigkeiten $\frac{m}{n}$ und $\frac{n-m}{n}$ wenig voneinander abweichen und daß jede von ihnen annähernd gleich 1/2 ist. Diese Gesetzmäßigkeit haben viele Forscher bemerkt, die lange Wurfserien mit einem Geldstück ausgeführt haben. BUFFON warf ein Geldstück 4040mal und erhielt 2048mal das Ergebnis Adler, also betrug die Häufigkeit des Adlers $\frac{m}{n} = 0{,}50693$. PEARSON, der 24000mal ein Geldstück warf, erhielt als Häufigkeit des Adlers 0,5005. Man sieht deutlich, daß die beobachtete Häufigkeit um die Zahl 0,5 schwankt.

Im Beispiel 1.1.2 kann man gleichfalls, wie langjährige Beobachtungen zeigen, eine gewisse Gesetzmäßigkeit feststellen. Mit dieser Frage werden wir uns im schon zitierten Beispiel 12.5.1 näher befassen.

Beispiel 1.1.3. Wir betrachten die Anzahl der Geburten von Jungen und Mädchen in Polen während der einzelnen Jahre des Zeitabschnitts von 1927 bis 1932. In keinem Einzelfall kann man das Geschlecht eines Neugeborenen voraussagen. Wir fassen das Geschlecht eines Neugeborenen als ein zufälliges Ereignis auf. Wenn man aber eine große Anzahl von Geburten beobachtet, wenn also eine Massenerscheinung vorliegt, dann kann man mit großer Genauigkeit voraussagen, wieviel Prozent der Neugeborenen Jungen und wieviel Prozent Mädchen sein werden.

In Tabelle 1.1.1 bezeichnen m bzw. k die Anzahl der Geburten von Jungen bzw. von Mädchen in den einzelnen Jahren. Die Häufigkeit der Geburten von Jungen bzw. von Mädchen ist mit p_1 bzw. p_2 bezeichnet, also

$$p_1 = \frac{m}{m+k}, \quad p_2 = \frac{k}{m+k}.$$

Wie man der Tabelle 1.1.1 entnimmt, schwanken die Werte p_1 um die Zahl 0,517 und die Werte p_2 um die Zahl 0,483.

Tabelle 1.1.1. *Die Geburten von Jungen und Mädchen*

Geburtsjahr	Geburtenanzahl		zusammen	Geburtenhäufigkeit	
	Jungen m	Mädchen k	$m+k$	Jungen p_1	Mädchen p_2
1927	496544	462189	958733	0,518	0,482
1928	513654	477339	990993	0,518	0,482
1929	514765	479336	994101	0,518	0,482
1930	528072	494739	1022811	0,516	0,484
1931	496986	467587	964573	0,515	0,485
1932	482431	452232	934663	0,516	0,484
zusammen bzw. Mittel	3032452	2833422	5865874	0,517	0,483

Beispiel 1.1.4. Wir werfen einen möglichst genau angefertigten Spielwürfel. Auf den Seiten des Würfels sind die Zahlen 1, 2, 3, 4, 5, 6 verzeichnet. Als Ergebnis eines Wurfes zeigt sich auf der oberen Seite des Würfels eine der Augenzahlen von 1 bis 6. Das Erscheinen irgendeiner bestimmten Zahl ist ein zufälliges Ereignis. Wenn wir eine lange Serie von Würfen ausführen und die Würfe zählen, in denen die Eins erscheint, dann bemerken wir, daß die Häufigkeit der Eins um 1/6 schwankt. Dasselbe trifft für das Erscheinen jeder anderen Zahl auf den Seiten des Würfels zu.

Diese beobachtete Gesetzmäßigkeit, eben die, daß die Häufigkeit des Eintreffens eines zufälligen Ereignisses bei einer großen Anzahl von Versuchen um einen festen Wert schwankt, liegt dem Wahrscheinlichkeitsbegriff zugrunde.

Zum Abschluß dieser einleitenden Bemerkungen betonen wir, daß wir die Wahrscheinlichkeitsrechnung nur auf solche Ereignisse anwenden, deren Häufigkeit wir unter gewissen Bedingungen entweder unmittelbar oder mittelbar beobachten oder über deren Eigenschaften wir mit Hilfe von logisch exakten Gedankengängen Aufschluß erhalten können.

1.2. Zufällige Ereignisse und Operationen mit zufälligen Ereignissen

A. Nun befassen wir uns mit der mathematischen Begriffsbestimmung eines zufälligen Ereignisses; die inhaltliche Bedeutung dieses Begriffs ist schon in 1.1 erläutert worden.

Der Grundbegriff der axiomatischen Theorie der Wahrscheinlichkeitsrechnung, zu deren Darstellung wir nun kommen, ist der Begriff der *Menge der Elementarereignisse*. Diese Menge bezeichnen wir mit dem Symbol E.

Bei jedem einzelnen Problem muß man feststellen, was in diesem Fall unter einem Elementarereignis zu verstehen ist, und die Menge E der Elementarereignisse angeben.

Beispiel 1.2.1. Nehmen wir an, daß wir uns beim Würfeln für die Häufigkeit einer geraden Augenzahl interessieren. Das Vorkommen einer einzelnen Zahl i, wobei $i = 1, 2, \ldots, 6$ ist, ist hier ein *Elementarereignis*, das wir mit e_i bezeichnen. Die Menge der Elementarereignisse enthält 6 Elemente.

In unserem Beispiel untersuchen wir das zufällige Ereignis A, das im Würfeln einer geraden Augenzahl, d. h. einer der Zahlen 2, 4, 6 besteht. Ein solches Ereignis schreiben wir in der Form

$$(e_2, e_4, e_6).$$

Damit meinen wir, daß das zufällige Ereignis (e_2, e_4, e_6) dann eintritt, wenn eines der Elementarereignisse e_2 oder e_4 oder e_6 eintritt.

Wenn wir das Erscheinen einer beliebigen Augenzahl außer der Eins beobachten wollten, dann hätten wir das fünfelementige Ereignis $(e_2, e_3, e_4, e_5, e_6)$ zu betrachten.

Unter der Menge Z der zufälligen Ereignisse verstehen wir hier die Menge aller Teilmengen der Menge E (also der Menge der Elementarereignisse).

Zur Menge Z gehören zunächst die einelementigen zufälligen Ereignisse

$$(e_1), (e_2), (e_3), (e_4), (e_5), (e_6);$$

dabei besteht z. B. das zufällige Ereignis (e_4) im Eintreffen des Elementarereignisses e_4, d. h. im Erscheinen der Zahl 4 auf der oberen Seite des Würfels.

Zur Menge Z gehören außer den 6 einelementigen zufälligen Ereignissen

$$(e_1), \ldots, (e_6)$$

noch die 15 zweielementigen Ereignisse

$$(e_1, e_2), \ldots, (e_5, e_6),$$

die 20 dreielementigen

$$(e_1, e_2, e_3), \ldots, (e_4, e_5, e_6),$$

die 15 vierelementigen

$$(e_1, e_2, e_3, e_4), \ldots, (e_3, e_4, e_5, e_6)$$

und die 6 fünfelementigen Ereignisse

$$(e_1, e_2, e_3, e_4, e_5), \ldots, (e_2, e_3, e_4, e_5, e_6).$$

Aber damit sind wir noch nicht zu Ende; denn die ganze Menge E ist ebenfalls ein Ereignis. Offenbar erhalten wir als Ergebnis eines Wurfes bestimmt eine der Zahlen 1, 2, 3, 4, 5, 6, also tritt bestimmt eines der Elementarereignisse der Menge E, mithin das Ereignis

$$(e_1, e_2, e_3, e_4, e_5, e_6)$$

ein. Im täglichen Leben betrachtet man ein Ereignis, dessen Eintreffen sicher ist, nicht als zufälliges Ereignis. Wir aber werden ein mit Sicherheit eintretendes Ereignis auch als zufälliges Ereignis ansehen und es zur Menge Z der zufälligen Ereignisse hinzuzählen.

Beim Würfeln betrachten wir endlich noch das Ereignis, daß eine größere Zahl als 6 vorkommt. Dieses Ereignis kommt unter den Elementen der Menge E nicht vor, ist also als Teilmenge der Menge E die leere Menge. Ein solches Ereignis ist natürlich unmöglich, und gewöhnlich betrachtet man es nicht als zufällig. Wir aber werden auch ein unmögliches Ereignis als zufälliges Ereignis ansehen, es als Teilmenge der Menge E zur Menge Z der zufälligen Ereignisse hinzuzählen und es mit dem Symbol (O) bezeichnen.

Zusammen mit dem unmöglichen und dem sicheren Ereignis enthält die Menge Z der zufälligen Ereignisse in unserem Beispiel 64 Ereignisse.

Hat allgemein die Menge der Elementarereignisse n Elemente, so enthält die Menge Z der zufälligen Ereignisse 2^n Ereignisse, nämlich:

1	unmögliches Ereignis (die leere Menge),
$\binom{n}{1}$,	einelementige Ereignisse,
$\binom{n}{2}$	zweielementige Ereignisse,
. . . .	. . . .
$\binom{n}{n-1}$	$(n-1)$-elementige Ereignisse,
1	sicheres Ereignis (die ganze Menge E).

B. Damit beenden wir die Betrachtungen, die sich auf das Beispiel 1.2.1 stützten. In diesem Beispiel war die Menge der Elementarereignisse endlich. In der Wahrscheinlichkeitsrechnung untersucht man aber auch Probleme, für die die Menge der Elementarereignisse abzählbar oder von der Mächtigkeit des Kontinuums ist. Im letzten Fall enthält die Menge Z der zufälligen Ereignisse nicht alle Ereignisse, d. h., sie enthält nicht alle Teilmengen der Menge E. Wir beschränken uns auf ein System von Teilmengen von E, die einen Borelschen[1]) Körper Z bilden. Der Leser findet die Definition eines solchen Mengensystems am Ende dieses Paragraphen, da das Buch auch einem Leser zugänglich sein soll, der die Mengenoperationen nicht kennt, während die Definition eines Borelschen Körpers die Kenntnisse solcher Operationen verlangt. Im Laufe der weiteren Darlegung wird auch erklärt, warum im allgemeinen Fall die Menge Z der zufälligen Ereignisse nicht als Menge aller Teilmengen von E definiert wird.

Wir kommen jetzt zur Definition eines zufälligen Ereignisses. In dieser Definition tritt der Begriff des Borelschen Körpers Z auf. Da dieser Begriff nicht vollständig präzisiert wurde, müssen wir in diesem Paragraphen auf die Definition des zufälligen Ereignisses zurückkommen (vgl. Definition 1.2.10).

Definition 1.2.1. Wir nennen jedes Element des Borelschen Körpers Z ein *zufälliges Ereignis*.

Der Körper Z ist also die Menge aller zufälligen Ereignisse.

Definition 1.2.2. Das Ereignis, das alle Elemente der Menge E enthält, nennen wir das *sichere Ereignis*.

Definition 1.2.3. Das Ereignis, das kein Element der Elementarereignismenge E enthält, nennen wir das *unmögliche Ereignis*.

---

[1]) Nach dem französischen Mathematiker Borel.

Das unmögliche Ereignis bezeichnen wir, wie schon gesagt, mit dem Symbol (O).

Definition 1.2.4. Wir sagen, daß *das Ereignis A in dem Ereignis B enthalten ist*, wenn jedes Elementarereignis, das zur Menge A gehört, auch in der Menge B vorhanden ist.

Wir schreiben

$$A \subset B$$

und lesen: A ist in B enthalten.

Diesen Sachverhalt soll Abb. 1.2.1 erläutern. Hier bedeuten das Quadrat E die Menge der Elementarereignisse und die Kreise A und B Teilmengen von E. Die Menge A ist in der Menge B enthalten.

Definition 1.2.4'. Wir sagen, daß *die Ereignisse A und B gleich sind*, wenn A in B und B in A enthalten ist.

Wir schreiben

$$A = B.$$

Wir setzen folgende Eigenschaften der Menge Z voraus:

Eigenschaft 1.2.1. Die Menge Z der zufälligen Ereignisse enthält als Element die Menge E der Elementarereignisse.

Eigenschaft 1.2.2. Die Menge Z der zufälligen Ereignisse enthält als Element die leere Menge (O).

Diese beiden Eigenschaften besagen, daß die Menge Z der zufälligen Ereignisse als Elemente das sichere und das unmögliche Ereignis enthält.

Definition 1.2.5. Wir sagen, *die Ereignisse A und B schließen einander aus*, wenn sie kein Elementarereignis gemeinsam haben.

Beispiel 1.2.2. Das zufällige Ereignis A bestehe darin, daß in einer Gruppe von n im Jahre 1950 geborenen Personen zwei das Jahr 2000 erleben, das zufällige Ereignis B darin, daß zwei oder mehr Personen der gleichen Gruppe bis zum Jahr 2000 leben. Die Ereignisse A und B schließen einander nicht aus.

Wenn wir dagegen das Ereignis B^* betrachten, welches darin besteht, daß nur eine einzige Person das Jahr 2000 erlebt, dann schließen die Ereignisse A und B^* einander aus.

Wir wollen dieses Beispiel näher untersuchen. Es kann geschehen, daß in der betrachteten Gruppe von n Personen nur eine Person das Jahr 2000 erlebt oder aber zwei, drei oder sogar n; es ist auch möglich, daß keine Person dieses Jahr erlebt. Die Menge E besteht hier aus $n + 1$ Elementarereignissen $e_0, e_1, \ldots, e_n$, wobei die Indizes $0, 1, \ldots, n$ die Anzahl der Personen bezeichnen, die bis zum Jahr 2000 leben. In diesem Beispiel enthält das zufällige Ereignis A ein einziges Element, und zwar das Element e_2. Das zufällige Ereignis B enthält $n - 1$ Elemente, nämlich die Elementarereignisse $e_2, e_3, \ldots, e_n$. Das gemeinsame Element der zufälligen Ereignisse A und B ist das Elementarereignis e_2, also schließen sich die beiden zufälligen Ereignisse nicht aus. Dagegen enthält das zufällige Ereignis B^* nur ein Element, nämlich das Element e_1. Die zufälligen Ereignisse A und B^* enthalten kein gemeinsames Element, schließen einander also aus.

C. Wir gehen nun zur Besprechung der Operationen über, die man für Ereignisse erklären kann. $A_1, A_2, \ldots$ sei eine endliche oder abzählbare Folge von Ereignissen.

Definition 1.2.6. Das Ereignis A, das alle Elementarereignisse und nur diese enthält, die zu irgendeinem der Ereignisse $A_1, A_2, \ldots$ gehören, nennen wir die *Summe* der Ereignisse $A_1, A_2, \ldots$.

Wir schreiben

$$A = A_1 \cup A_2 \cup \ldots \quad \text{oder} \quad A = A_1 + A_2 + \cdots \quad \text{oder} \quad A = \sum_i A_i$$

und lesen: A gleich A_1 oder A_2 oder ....

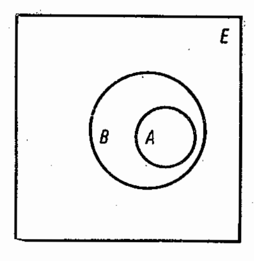


Abb. 1.2.1

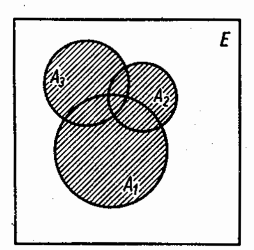


Abb. 1.2.2

Wir wollen die Summe von Ereignissen an einer Zeichnung (Abb. 1.2.2) erläutern. Hier bedeutet das Quadrat E die Menge der Elementarereignisse, während die Kreise A_1, A_2, A_3 drei Ereignisse darstellen; das schraffierte Gebiet stellt die Summe $A_1 + A_2 + A_3$ dar.

Die hier gegebene Definition der Summe der Ereignisse entspricht der mengentheoretischen Vereinigung der Teilmengen $A_1, A_2, \ldots$ in der Menge E der Elementarereignisse. Die Summe der Ereignisse tritt genau dann ein, wenn wenigstens eines dieser Ereignisse eintritt.

Hier erhebt sich die wichtige Frage, ob die Summe beliebiger endlich oder abzählbar vieler zufälliger Ereignisse zur Menge Z gehört, d. h., ob sie wieder ein zufälliges Ereignis ist.

Soll die Frage bejaht werden, so muß die Menge Z der zufälligen Ereignisse die folgende Eigenschaft besitzen:

Eigenschaft 1.2.3. Gehören die endlich oder abzählbar vielen Ereignisse $A_1, A_2, \ldots$ zur Menge Z, so gehört ihre Summe ebenfalls zu Z.

Man bestätigt leicht, daß für ein beliebiges Ereignis A folgende Gleichungen gelten:

$$A \cup A = A, \quad A \cup E = E, \quad A \cup (O) = A.$$

Wir wollen als Beispiel für Beweise dieser Art die Beziehung $A \cup A = A$ beweisen. Das Elementarereignis e gehöre zu A. Dann gehört es auch zu $A \cup A$, woraus die Relation $A \subset (A \cup A)$ folgt. Gehört andererseits e zu $A \cup A$, so auch zu A, woraus $(A \cup A) \subset A$ folgt. Nach Definition 1.2.4′ gilt also $A = A \cup A$.

Definition 1.2.7. Das Ereignis A, welches die Elementarereignisse und nur diese enthält, die zum Ereignis A_1, aber nicht zum Ereignis A_2 gehören, nennen wir die *Differenz* der Ereignisse A_1 und A_2.

Wir schreiben

$$A = A_1 - A_2.$$

Die Differenz von Ereignissen wird an Abb. 1.2.3 erläutert. Hier bedeutet das Quadrat E die Menge der Elementarereignisse, die Kreise A_1 und A_2 sind zwei Ereignisse. Das schraffierte Gebiet stellt die Differenz $A_1 - A_2$ dar.

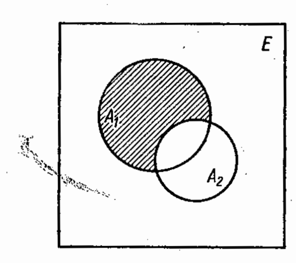


Abb. 1.2.3

Die Differenz $A_1 - A_2$ der Ereignisse tritt genau dann ein, wenn das Ereignis A_1, nicht aber das Ereignis A_2 eintritt.

Schließen die Ereignisse A_1 und A_2 einander aus, so stimmt die Differenz $A_1 - A_2$ mit dem Ereignis A_1 überein.

Wie oben fordern wir die folgende Eigenschaft der Menge Z der zufälligen Ereignisse:

Eigenschaft 1.2.4. Gehören die Ereignisse A_1 und A_2 zur Menge Z, so gehört auch ihre Differenz zu Z.

Beispiel 1.2.3. Wir betrachten eine bestimmte Gruppe von Familien. Die Anzahl der Kinder einer dieser Familien ist ein zufälliges Ereignis. Das Ereignis A bestehe darin, daß eine zufällig [1]) gewählte Familie genau ein Kind hat, das Ereignis B darin, daß diese Familie wenigstens ein Kind hat. Die Summe $A + B$ der Ereignisse A und B besteht dann darin, daß die zufällig gewählte Familie wenigstens ein Kind hat.

Die höchste Kinderzahl in einer Familie der Gruppe sei n. Die Menge der Elementarereignisse besteht dann aus $n + 1$ Elementen, welche wir wie in Beispiel 1.2.2 mit den Symbolen $e_0, e_1, e_2, \ldots, e_n$ bezeichnen wollen. Da das Ereignis A nur das Elementarereignis e_1 und das Ereignis B die Elementarereignisse $e_1, \ldots, e_n$ enthält, besteht ihre Summe aus den Elementarereignissen $e_1, e_2, \ldots, e_n$. Die Differenz $A - B$ der Ereignisse ist hier selbstverständlich ein unmögliches Ereignis, da es kein Elementarereignis gibt, das zu A, nicht aber zu B gehört. Die Differenz $B - A$ dagegen enthält die Elemente $e_2, e_3, \ldots, e_n$ und ist das Ereignis, daß die willkürlich gewählte Familie mehr als ein Kind besitzt.

Definition 1.2.8. Das Ereignis A, das alle Elementarereignisse und nur diese enthält, die in allen Ereignissen $A_1, A_2, \ldots$ enthalten sind, nennen wir das *Produkt* dieser Ereignisse.

[1]) Wie man eine zufällige Auswahl trifft, wird später erklärt.

Wir schreiben

$$A = A_1 \cap A_2 \cap \ldots \quad \text{oder} \quad A = A_1 A_2 \cdots \quad \text{oder} \quad A = \prod_i A_i$$

und lesen: A gleich A_1 und A_2 und ....

Das Produkt von Ereignissen wollen wir an Abb. 1.2.4 erläutern. Hier bedeutet das Quadrat E die Menge der Elementarereignisse, und die Kreise A_1, A_2, A_3 sind drei Ereignisse; das schraffierte Gebiet stellt das Produkt $A_1 A_2 A_3$ dar.

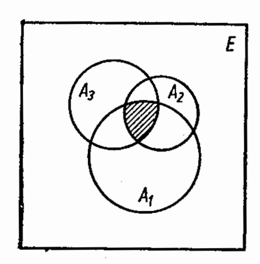


Abb. 1.2.4

Die hier gegebene Definition des Produkts von Ereignissen entspricht dem Durchschnitt der Teilmengen $A_1, A_2, \ldots$ in der Menge E der Elementarereignisse. Das Produkt A von Ereignissen A_i tritt genau dann ein, wenn alle Ereignisse A_i eintreten.

Wir fordern die folgende Eigenschaft der Menge Z:

Eigenschaft 1.2.5. Gehören die endlich oder abzählbar vielen Ereignisse $A_1, A_2, \ldots$ zur Menge Z, so gehört ihr Produkt ebenfalls zu Z.

Man bestätigt leicht für ein beliebiges Ereignis A die Gleichungen

$$A \cap A = A, \quad A \cap E = A, \quad A \cap (O) = (O).$$

Beispiel 1.2.4. Das Ereignis A bestehe darin, daß eine zufällig gewählte Bauernwirtschaft wenigstens ein Pferd und nicht weniger als einen Pflug hat, wobei die höchstmögliche Anzahl der Pferde sowie der Pflüge gleich Zwei sein soll. Das Ereignis B bestehe darin, daß in der Wirtschaft genau ein Pferd und nicht mehr als ein Pflug vorhanden ist. Wir suchen das Produkt der Ereignisse A und B.

Die Menge der Elementarereignisse besteht hier aus 9 Elementen, die wir mit den Symbolen

$$e_{00}, e_{01}, e_{02}, e_{10}, e_{11}, e_{12}, e_{20}, e_{21}, e_{22}$$

bezeichnen, wobei der erste Index die Anzahl der Pferde und der zweite die Anzahl der Pflüge angibt, die in der Wirtschaft vorhanden sind.

Das Ereignis A enthält vier Elementarereignisse $e_{11}, e_{12}, e_{21}, e_{22}$, das Ereignis B zwei, nämlich e_{10} und e_{11}. Das Produkt $A \cap B$ enthält ein Elementarereignis e_{11} und bedeutet, daß in der Wirtschaft genau ein Pferd und genau ein Pflug vorhanden sind.

Definition 1.2.9. Wir nennen die Differenz $E - A$ das zu A *komplementäre Ereignis* und bezeichnen es mit $\bar{A}$.

Abb. 1.2.5 veranschaulicht das zu A komplementäre Ereignis. Hier bedeuten E die Menge der Elementarereignisse und der Kreis A ein Ereignis, während das schraffierte Gebiet das zu A komplementäre Ereignis $\bar{A}$ darstellt.

Die obige Definition kann man auch so formulieren:

Das Ereignis $\bar{A}$, daß das Ereignis A nicht eintritt, nennen wir das zu A *komplementäre Ereignis.*

Nach den Eigenschaften 1.2.1 und 1.2.4 der Menge Z der zufälligen Ereignisse ist das Ereignis $\bar{A}$ ebenfalls ein zufälliges Ereignis.

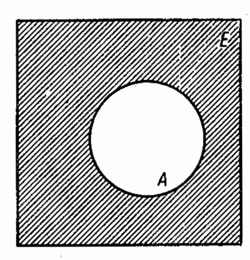


Abb. 1.2.5

Beispiel 1.2.5. Wir haben einen Satz Glühlampen und interessieren uns für deren Brenndauer t. Dazu setzen wir einen Wert t_1 fest und bezeichnen eine Glühlampe, die in einer kürzeren Zeit als t_1 durchbrennt, als fehlerhaft. Nun greifen wir eine Glühlampe willkürlich heraus. Das Ereignis A bestehe im Herausgreifen einer fehlerhaften Glühlampe. Dann ist das Herausgreifen einer fehlerlosen Glühlampe, die wenigstens während t_1 Zeiteinheiten brennt, das zu A komplementäre Ereignis $\bar{A}$.

Wir bringen jetzt die schon angekündigte Definition (vgl. Anhang) eines Borelschen Mengenkörpers von zufälligen Ereignissen.

Definition 1.2.10. Eine Menge Z von Teilmengen einer bestimmten Menge E von Elementarereignissen mit den Eigenschaften 1.2.1 bis 1.2.5 nennen wir einen *Borelschen Mengenkörper* und dessen Elemente *zufällige Ereignisse.*

In Zukunft wollen wir uns ausschließlich mit zufälligen Ereignissen befassen und werden deshalb statt „zufälliges Ereignis" oft nur kurz „Ereignis" sagen.

Die folgenden Bezeichnungen sollen dazu dienen, die Formulierungen und Beweise einiger Sätze im Laufe der weiteren Darstellung zu vereinfachen.

Definition 1.2.11. Wir nennen eine Folge $\{A_n\}$ $(n = 1, 2, \ldots)$ von Ereignissen *absteigend*, wenn für jedes n die Beziehung

$$A_{n+1} \subset A_n$$

gilt. Das Produkt A einer absteigenden Folge $\{A_n\}$ nennen wir den *Limes* dieser Folge und schreiben

$$A = \prod_{n \geqq 1} A_n = \lim_{n \to \infty} A_n .$$

Definition 1.2.12. Wir nennen eine Folge $\{A_n\}$ $(n = 1, 2, \ldots)$ von Ereignissen *aufsteigend*, wenn für jedes n die Beziehung

$$A_{n+1} \supset A_n$$

gilt. Die Summe A einer aufsteigenden Folge $\{A_n\}$ nennen wir den *Limes* dieser Folge und schreiben

$$A = \sum_{n \geq 1} A_n = \lim_{n \to \infty} A_n.$$

1.3. Das Axiomensystem der Wahrscheinlichkeitsrechnung

A. In der Umgangssprache benutzt man den Wahrscheinlichkeitsbegriff ohne genaue Bestimmung seiner Bedeutung. Die Wahrscheinlichkeitsrechnung als ein Teil der mathematischen Wissenschaft aber muß diesen Begriff genau festlegen. Das soll durch ein System von Axiomen geschehen, die einige grundlegende Eigenschaften dieses Begriffes angeben, d. h. durch eine Axiomatisierung der Wahrscheinlichkeitsrechnung.[1]) Die weiteren Eigenschaften der Wahrscheinlichkeiten erhalten wir als Sätze, die aus diesen Axiomen hergeleitet werden.

Der exakte Begriff des zufälligen Ereignisses, dessen Bedeutung im vorhergehenden Paragraphen erklärt wurde, ist die mathematische Formulierung dessen, was man in der Umgangssprache gewöhnlich unter einem zufälligen Ereignis versteht. Das Axiomensystem, das wir jetzt formulieren wollen und das den Begriff der Wahrscheinlichkeit eines zufälligen Ereignisses präzisiert, ist die mathematische Fassung der Gesetzmäßigkeiten, die für die Häufigkeiten der zufälligen Ereignisse beobachtet werden konnten. Diese Gesetzmäßigkeiten treten im Ergebnis von langen Versuchsreihen auf, die unter denselben Bedingungen ausgeführt werden.

Wie schon bemerkt wurde (vgl. 1.1), hat man beobachtet, daß die Häufigkeit des Eintretens eines zufälligen Ereignisses um einen festen Wert schwankt, wenn die Versuchsserie genügend lang ist. Diese für die Häufigkeit von zufälligen Ereignissen beobachtete Gesetzmäßigkeit und die Tatsache, daß die Häufigkeit ein nichtnegativer Bruch ist, der höchstens gleich Eins ist, legt die Annahme des folgenden Axioms nahe. Dabei sei eine bestimmte Menge E von Elementarereignissen und ein Borelscher Körper Z von Teilmengen von E gegeben.

---

[1]) Der axiomatischen Fassung der Wahrscheinlichkeitsrechnung sind viele Arbeiten gewidmet. Wir nennen hier die Arbeiten von Bernstein [1], Łomnicki [1], Rényi [4], Steinhaus [1] und das Buch von Mazurkiewicz [1]. Das axiomatische System, das wir in diesem Paragraphen angeben, stammt von Kolmogoroff [7]. (Die Zahlen in den Klammern beziehen sich auf die Nummer der zitierten Arbeit im Literaturverzeichnis, das sich am Ende des Buches befindet.) Die Grundbegriffe der Wahrscheinlichkeitsrechnung sind auch in dem bekannten Werk von Laplace [1] sowie in den Arbeiten von Hausdorff [1], von Mises [1], [2], Jeffreys [1] und Barankin [2] untersucht worden.

Axiom I. *Jedem zufälligen Ereignis A entspricht eine bestimmte Zahl $P(A)$, seine Wahrscheinlichkeit, welche die Ungleichung*

$$0 \leqq P(A) \leqq 1$$

erfüllt.

Das folgende einfache Beispiel führt intuitiv zur Formulierung eines weiteren Axioms.

Beispiel 1.3.1. In einer Urne befinden sich nur schwarze Kugeln. Das Experiment bestehe im Herausgreifen einer Kugel aus der Urne. Es bezeichne m/n wie bisher die Häufigkeit einer schwarzen Kugel. Dann ist klar, daß hier immer $m/n = 1$ ist. Die Auswahl einer schwarzen Kugel ist in diesem Fall ein sicheres Ereignis, und seine Häufigkeit ist, wie wir sehen, gleich Eins.

Bei der Formulierung des folgenden Axioms berücksichtigen wir diese Eigenschaft des sicheren Ereignisses.

Axiom II. *Die Wahrscheinlichkeit des sicheren Ereignisses ist gleich Eins.*

Diese Eigenschaft läßt sich durch die Formel

$$P(E) = 1$$

ausdrücken.

Wir werden in 2.3 sehen, daß die Umkehrung des Axioms II nicht gilt: *Wenn die Wahrscheinlichkeit eines zufälligen Ereignisses A gleich Eins ist, also $P(A) = 1$, dann braucht die Menge A nicht alle Elementarereignisse der Menge E zu enthalten.*

Wir haben schon gesehen, daß die Häufigkeit einer Sechs beim Würfeln um die Zahl $1/6$ schwankt. Dasselbe gilt für die Häufigkeit der Zwei. Offenbar schließen diese beiden Ereignisse einander aus, und man beobachtet, daß die Häufigkeit einer Zwei oder Sechs, d. h. die Häufigkeit der Summe dieser Ereignisse, gleich der Summe der Häufigkeiten ist und um die Zahl $1/6 + 1/6$, d. h. um $1/3$ schwankt.

Der Versuch lehrt, daß in einem Kartenspiel mit 52 Karten (4 Farben zu je 13 Karten) die Häufigkeit irgendeines der 4 Asse, wenn man sehr oft eine Karte zieht, ungefähr $4/52$ beträgt, die Häufigkeit der 13 Pique-Karten ungefähr $13/52$; dagegen schwankt die Häufigkeit der Summe „As oder Pique" nicht um die Zahl $4/52 + 13/52 = 17/52$, sondern um die Zahl $16/52$. Das erklärt sich daraus, daß die zufälligen Ereignisse „As" oder „Pique" einander nicht ausschließen (man kann das Pique-As ziehen); die Häufigkeit der Summe „As oder Pique" ist also nicht gleich der Summe der Häufigkeiten der Ereignisse „As" und „Pique". Bei der Formulierung des folgenden letzten Axioms berücksichtigen wir diese Eigenschaft der Häufigkeit von Summen.

Axiom III. *Die Wahrscheinlichkeit einer Summe von endlich oder abzählbar vielen zufälligen Ereignissen, die einander paarweise ausschließen, ist gleich der Summe der Wahrscheinlichkeiten dieser Ereignisse.*

Unter Anwendung dieses Axioms kann man die Wahrscheinlichkeit eines zufälligen Ereignisses, das aus endlich oder abzählbar vielen Elementarereignissen e_k besteht, wobei $(e_k) \in Z$ $(k = 1, 2, \ldots)$, durch die Summe von Wahrscheinlichkeiten von einelementigen Ereignissen ausdrücken:

$$P(e_1, e_2, \ldots) = P(e_1) + P(e_2) + \cdots.$$

Wenn wir allgemein eine endliche oder abzählbare Folge von zufälligen Ereignissen $\{A_k\}$ $(k = 1, 2, \ldots)$ haben, wobei diese Ereignisse einander paarweise ausschließen, dann gilt nach Axiom III die Formel

$$P\Big(\sum_k A_k\Big) = \sum_k P(A_k). \tag{1.3.1}$$

Die durch das Axiom III angegebene Eigenschaft nennen wir die *vollständige Additivität der Wahrscheinlichkeit.*[1])

Axiom III bezieht sich nur auf Summen von Ereignissen, die einander paarweise ausschließen. Jetzt seien A und B beliebige zufällige Ereignisse, die einander nicht notwendig ausschließen. Wir wollen die Wahrscheinlichkeit der Summe dieser Ereignisse bestimmen.

Es gilt

$$A \cup B = A \cup (B - AB),$$
$$B = AB \cup (B - AB).$$

Die rechten Seiten dieser Gleichungen sind Summen von Ereignissen, die einander ausschließen, also erhält man nach Axiom III die Beziehungen

$$P(A \cup B) = P(A) + P(B - AB),$$
$$P(B) = P(AB) + P(B - AB).$$

Aus beiden Gleichungen erhalten wir für die Wahrscheinlichkeit der Summe der Ereignisse A und B die Formel

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1.3.2}$$

Es seien $A_1, A_2, \ldots, A_n$ $(n \geqq 3)$ beliebige zufällige Ereignisse. Dann läßt sich leicht die folgende (von Poincaré stammende) Formel herleiten:

$$P\left(\sum_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k) - \sum_{\substack{k_1,k_2=1\\k_1<k_2}}^{n} P(A_{k_1}A_{k_2}) + \sum_{\substack{k_1,k_2,k_3=1\\k_1<k_2<k_3}}^{n} P(A_{k_1}A_{k_2}A_{k_3}) + \cdots + (-1)^{n-1} P(A_1 \cdots A_n). \tag{1.3.2'}$$

[1]) Man sagt auch, die Wahrscheinlichkeit $P(A)$, die die angegebenen Axiome I bis III erfüllt, ist ein normiertes, nichtnegatives, vollständig additives Maß auf dem Borelschen Mengenkörper Z von Teilmengen der Menge E.

B. Wir betrachten endlich oder abzählbar viele zufällige Ereignisse A_k ($k = 1, 2, \ldots$). Wenn jedes Elementarereignis der Menge E wenigstens zu einem der zufälligen Ereignisse $A_1, A_2, \ldots$ gehört, dann sagen wir, daß diese zufälligen Ereignisse die Menge E *ausschöpfen* oder *überdecken*. Die Summe $\sum_k A_k$ enthält alle Elementarereignisse von E, sie ist also das sichere Ereignis. Gemäß Axiom II erhalten wir:

Satz 1.3.1. *Wenn endlich oder abzählbar viele zufällige Ereignisse $A_1, A_2, \ldots$ die Menge der Elementarereignisse E ausschöpfen, dann gilt*

$$P\left(\sum_k A_k\right) = 1. \tag{1.3.3}$$

Beispiel 1.3.2. Die Menge aller nichtnegativen ganzen Zahlen sei die Menge E der Elementarereignisse. Das Ereignis, das im Auftreten der Zahl n besteht ($n = 0, 1, 2, \ldots$), bezeichnen wir mit (e_n). Nun nehmen wir an, daß

$$P(e_n) = \frac{c}{n!}$$

ist; dabei ist c ein Proportionalitätsfaktor. Aus Satz 1.3.1 und Axiom III folgt

$$P\left(\sum_{n=0}^{\infty} e_n\right) = c \sum_{n=0}^{\infty} \frac{1}{n!}.$$

Nun ist aber

$$P\left(\sum_{n=0}^{\infty} e_n\right) = 1 \quad \text{und} \quad \sum_{n=0}^{\infty} \frac{1}{n!} = e,$$

wobei e die Basis der natürlichen Logarithmen ist. Wir erhalten also $1 = ce$ und folglich $c = e^{-1}$.

Wie es sich in den weiteren Kapiteln zeigen wird, haben wir es in diesem Beispiel mit einem Spezialfall der Poissonverteilung zu tun, die häufig in Anwendungen vorkommt.

Wir beweisen den folgenden Satz:

Satz 1.3.2. *Die Summe der Wahrscheinlichkeiten des zufälligen Ereignisses A und des zu A komplementären Ereignisses $\bar{A}$ ist gleich Eins.*

Beweis. Aus der Definition des Ereignisses $\bar{A}$ folgt, daß die Summe $A \cup \bar{A}$ der Ereignisse A und $\bar{A}$ ein sicheres Ereignis ist. Wir erhalten also nach Axiom II

$$P(A \cup \bar{A}) = 1.$$

Die Ereignisse A und $\bar{A}$ schließen aber einander aus; also ist nach Axiom III

$$P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

und schließlich

$$P(A) + P(\bar{A}) = 1. \tag{1.3.4}$$

A sei ein unmögliches Ereignis. Wir beweisen folgenden

Satz 1.3.3. *Die Wahrscheinlichkeit des unmöglichen Ereignisses ist gleich Null.*

Beweis. Für jedes zufällige Ereignis A gilt die Gleichung

$$A \cup E = E.$$

Wenn A das unmögliche Ereignis ist (also kein Elementarereignis enthält), dann schließen A und E einander aus, da sie kein gemeinsames Element besitzen. Durch Anwendung von Axiom III erhalten wir

$$P(A) + P(E) = P(E).$$

Daraus folgt

$$P(A) = 0.$$

Wir werden in 2.3 sehen, daß die Umkehrung dieses Satzes falsch ist; *wenn die Wahrscheinlichkeit eines Ereignisses gleich Null ist, folgt daraus noch nicht, daß es das unmögliche Ereignis ist.*

C. Die beiden folgenden Sätze finden vielfach Anwendung:

Satz 1.3.4. *Es sei $\{A_n\}$ $(n = 1, 2, \ldots)$ eine absteigende Folge von Ereignissen und A das Produkt der Ereignisse A_n. Dann gilt die Beziehung*

$$P(A) = \lim_{n\to\infty} P(A_n). \tag{1.3.5}$$

Beweis. Da die Folge $\{A_n\}$ absteigend ist, gilt für jedes n die Gleichung

$$A_n = \sum_{k=n}^{\infty} A_k \bar{A}_{k+1} + A.$$

Aus der Formel (1.3.2) folgt weiter

$$P(A_n) = P\left(\sum_{k=n}^{\infty} A_k \bar{A}_{k+1}\right) + P(A) - P\left(A \sum_{k=n}^{\infty} A_k \bar{A}_{k+1}\right). \tag{1.3.6}$$

Nun ist

$$A \sum_{k=n}^{\infty} A_k \bar{A}_{k+1} = \sum_{k=n}^{\infty} A A_k \bar{A}_{k+1};$$

das Ereignis

$$A A_k \bar{A}_{k+1}$$

ist jedoch für jedes k ein unmögliches Ereignis, also ist

$$P(A A_k \bar{A}_{k+1}) = 0.$$

Nach Axiom III erhalten wir dann

$$P\left(\sum_{k=n}^{\infty} A A_k \bar{A}_{k+1}\right) = 0.$$

Da sich die Ereignisse in der Summe auf der rechten Seite der Formel (1.3.6) paarweise ausschließen, haben wir

$$P(A_n) = \sum_{k=n}^{\infty} P(A_k \bar{A}_{k+1}) + P(A). \tag{1.3.7}$$

Andererseits ist die Reihe

$$\sum_{k=1}^{\infty} P(A_k \bar{A}_{k+1})$$

konvergent; denn sie ist eine Summe von nichtnegativen Summanden, die höchstens gleich Eins ist. Daraus folgt, daß für $n \to \infty$ die Summe in Formel (1.3.7) gegen Null strebt. Also erhält man

$$\lim_{n \to \infty} P(A_n) = P(A).$$

Satz 1.3.5. *Es sei $\{A_n\}$ $(n = 1, 2, \ldots)$ eine aufsteigende Folge von Ereignissen und A die Summe der Ereignisse A_n. Dann gilt die Beziehung*

$$P(A) = \lim_{n \to \infty} P(A_n). \tag{1.3.8}$$

Beweis. Wir betrachten die Folge $\{\bar{A}_n\}$ der zu A_n komplementären Ereignisse. Da $\{A_n\}$ eine aufsteigende Folge ist, muß die Folge $\{\bar{A}_n\}$ absteigend sein. Es sei $\bar{A}$ das Produkt der Ereignisse $\bar{A}_n$. Aus Satz 1.3.4 folgt

$$P(\bar{A}) = \lim_{n \to \infty} P(\bar{A}_n)$$

und daraus

$$P(A) = 1 - P(\bar{A}) = 1 - \lim_{n \to \infty} P(\bar{A}_n) = 1 - \lim_{n \to \infty} [1 - P(A_n)]$$
$$= \lim_{n \to \infty} P(A_n).$$

Damit ist der Satz bewiesen.

Wir wollen noch einen einfachen Satz angeben:

Satz 1.3.6. *Gilt für die zufälligen Ereignisse A und B die Beziehung*

$$A \subset B,$$

dann ist.

$$P(A) \leqq P(B).$$

Beweis. Wir schreiben

$$B = A + (B - A).$$

Die Ereignisse A und $B - A$ schließen einander aus, also ist nach Axiom III

$$P(B) = P(A) + P(B - A).$$

Da $P(B - A) \geqq 0$ ist, erhalten wir $P(B) \geqq P(A)$.

1.4. Anwendungen der Kombinatorik zur Berechnung von Wahrscheinlichkeiten

Bei einigen Problemen kann man die Wahrscheinlichkeiten mit Hilfe von Formeln aus der Kombinatorik berechnen.[1]) Wir wollen das an Beispielen zeigen.

Beispiel 1.4.1. In einer Urne befinden sich 5 verschiedenfarbige Kugeln. Wir nehmen an, daß die Wahrscheinlichkeit, gezogen zu werden, für jede Kugel dieselbe, und zwar gleich p ist.

Die Menge E besteht hier aus 5 Elementarereignissen, und nach unserer Annahme hat jedes von ihnen die gleiche Wahrscheinlichkeit. Auf Grund von Satz 1.3.1 erhalten wir $5p = 1$, also $p = 1/5$.

Beispiel 1.4.2. In einer Urne befinden sich 9 Papierstreifen gleicher Größe. Die Streifen seien so mit den Zahlen 1 bis 9 versehen, daß nicht zwei Streifen gleich bezeichnet sind.

Die Menge E besteht hier aus 9 Elementarereignissen. Das zufällige Ereignis, das im Ziehen eines Streifens mit einer geraden Zahl besteht, bezeichnen wir mit A. Wie groß ist seine Wahrscheinlichkeit?

Ähnlich wie vorher setzen wir voraus, daß die Wahrscheinlichkeit, gezogen zu werden, für jeden einzelnen Streifen dieselbe, also gleich 1/9 ist. Einen Streifen mit einer geraden Zahl erhalten wir dann, wenn wir einen Streifen ziehen, der mit einer der Zahlen 2, 4, 6, 8 versehen ist. Nach Axiom III ist die gesuchte Wahrscheinlichkeit

$$P(A) = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{4}{9}.$$

Wollen wir für dieses Beispiel die Wahrscheinlichkeit eines mit einer ungeraden Ziffer versehenen Streifens berechnen, so müssen wir nur beachten, daß dieses zufällige Ereignis zum Ereignis A komplementär ist, und wir erhalten, wenn wir es mit $\bar{A}$ bezeichnen, durch Anwendung von Satz 1.3.2

$$P(\bar{A}) = 1 - P(A) = \frac{5}{9}.$$

---

[1]) Diese Formeln findet man z. B. im Buch von MOSTOWSKI und STARK [1], S. 12, sowie in FLACHSMEYER [1].

Beispiel 1.4.3. Wir werfen eine Münze dreimal. Wie groß ist die Wahrscheinlichkeit dafür, daß der Adler zweimal vorkommt? Für die Ergebnisse von drei Würfen gibt es $2^3 = 8$ verschiedene Fälle. Bezeichnen wir das Vorkommen des Adlers mit A und das der Zahl mit K, so sind folgende Anordnungen möglich:

$$AAA, AAK, AKA, KAA, AKK, KAK, KKA, KKK.$$

Diese Anordnungen bilden die Menge E. Jede dieser Anordnungen sehen wir als Elementarereignis an und setzen voraus, daß sie mit gleicher Wahrscheinlichkeit auftreten. Dann ist die Wahrscheinlichkeit einer beliebigen Anordnung gleich $1/2^3$. Aus der obigen Zusammenstellung sehen wir, daß das Ereignis „zweimal Adler" in drei Elementarereignissen vorkommt (AAK, AKA, KAA). Also ist auf Grund von Axiom III die gesuchte Wahrscheinlichkeit gleich 3/8.

Wenn wir in unserem Beispiel n Würfe an Stelle von drei betrachten und die Wahrscheinlichkeit suchen, m-mal den Adler zu erhalten, so müssen wir folgendermaßen schließen:

Bei n Würfen gibt es 2^n Möglichkeiten für die Anordnung der Ergebnisse. Anordnungen, in denen der Adler m-mal vorkommt, gibt es so viele, wie es Kombinationen zu je m Elementen aus n Elementen gibt:

$$\binom{n}{m} = \frac{n!}{m!\,(n-m)!}.$$

Ist jede Anordnung, die man im Ergebnis von n Würfen erhält, gleich wahrscheinlich, dann beträgt die gesuchte Wahrscheinlichkeit

$$\frac{n!}{2^n m!\,(n-m)!}. \tag{1.4.1}$$

Beispiel 1.4.4. Wir wollen die Wahrscheinlichkeit dafür berechnen, daß bei drei Würfen der Adler wenigstens zweimal vorkommt.

Das betrachtete zufällige Ereignis tritt dann ein, wenn bei drei Würfen der Adler zweimal oder dreimal vorkommt. Nach Formel (1.4.1) beträgt die Wahrscheinlichkeit dafür, daß der Adler dreimal vorkommt,

$$\frac{3!}{2^3 3!\,0!} = \frac{1}{8},$$

während die Wahrscheinlichkeit, daß der Adler zweimal vorkommt, wie wir schon wissen, 3/8 ist. Auf Grund von Axiom III ist also die gesuchte Wahrscheinlichkeit

$$\frac{1}{8} + \frac{3}{8} = \frac{1}{2}.$$

In den Beispielen 1.4.1 bis 1.4.4 wurde vorausgesetzt, daß alle Elementarereignisse gleich wahrscheinlich sind. Diese Voraussetzung war in den angegebenen Beispielen selbstverständlich, sie braucht aber nicht immer erfüllt zu sein.

1.5. Die bedingte Wahrscheinlichkeit

A. Beispiel 1.5.1. Markoff [4] untersuchte die Wahrscheinlichkeit für das Auftreten der folgenden Buchstabenpaare in der russischen Sprache:

Selbstlaut nach Selbstlaut

Selbstlaut nach Mitlaut.

Zu diesem Zweck zählte er die entsprechenden Buchstabenpaare und berechnete deren Häufigkeit in Puschkins Versroman „Eugen Onegin", und zwar in einem Textabschnitt, der 20000 Buchstaben umfaßte. Diese beobachteten Häufigkeiten betrachtete er als Wahrscheinlichkeiten.[1] Die Untersuchung ergab Folgendes: Es waren an Selbstlauten 8638 vorhanden, dabei kam das Paar „Selbstlaut nach Selbstlaut" 1104mal vor.

Wir wollen dieses Beispiel näher betrachten. Mit a sei ein Selbstlaut, mit b ein Mitlaut bezeichnet. Als Elementarereignisse sehen wir die Paare aa, ba, ab, bb an, also als Menge E die Menge (aa, ba, ab, bb).

Das Ereignis B bestehe im Auftreten eines Buchstabenpaares, das an zweiter Stelle einen Selbstlaut hat. Das Ereignis B kann man in Form der Menge (aa, ba) schreiben. Man weiß, daß Selbstlaute (a) im ganzen 8638mal vorkommen. Den Selbstlauten geht entweder ein Selbstlaut voraus (die Paare aa) oder ein Mitlaut (die Paare ba).

Da am Anfang des untersuchten Textes:

„Мой дядя самых честных правил..."

kein Selbstlaut steht, kommt also das Ereignis B genau 8638mal vor. Deshalb ist

$$P(B) = \frac{8\,638}{20\,000} = 0{,}432.$$

Ferner bestehe das Ereignis A im Vorkommen eines Buchstabenpaares, das an erster Stelle einen Selbstlaut hat. Das Ereignis A kann man in Form der Menge (aa, ab) schreiben.

Die Frage nach der Häufigkeit eines Selbstlautes unter der Bedingung, daß ihm ein Selbstlaut folgt, kann man auch folgendermaßen formulieren:

Wie groß ist die Wahrscheinlichkeit des zufälligen Ereignisses A unter der Bedingung, daß das zufällige Ereignis B eingetreten ist? Es handelt sich hier nicht um die Wahrscheinlichkeit des Ereignisses A, betrachtet in der ganzen Menge E der Elementarereignisse, sondern um die bedingte Wahrscheinlichkeit. Dieser entspricht die bedingte Häufigkeit des Ereignisses A, und zwar unter der Bedingung, daß das Ereignis B eintrat. Es handelt sich also um die Wahrscheinlichkeit des Ereignisses A in der Menge (aa, ba), die hier als Menge der Elementarereignisse betrachtet wird.

[1] Im zweiten Teil dieses Buches geben wir Methoden an, die es erlauben, die Berechtigung derartiger Hypothesen nachzuprüfen.

In unserem Beispiel geht es um die Wahrscheinlichkeit von (aa). Die Untersuchung erwies, daß dieses Ereignis im ganzen 1104mal vorkommt; da aber das Ereignis B genau 8638mal vorkommt, beträgt die gesuchte bedingte Wahrscheinlichkeit

$$\frac{1104}{8638} = 0{,}128.$$

B. Es sei B ein zufälliges Ereignis in einer gewissen Menge E von Elementarereignissen. Die Menge B ist also ein Element des Borelschen Mengenkörpers Z, der aus Teilmengen der Menge E besteht. Es sei $P(B) > 0$. Wir betrachten B als neue Menge von Elementarereignissen und bezeichnen mit Z' den Borelschen Mengenkörper, der aus allen zum Körper Z gehörenden Teilmengen von B gebildet ist.

Wir betrachten ein beliebiges, zum Körper Z gehörendes Ereignis A. Es kann in einem besonderen Fall zum Körper Z' gehören, nämlich dann, wenn es eine Teilmenge von B ist. Enthält dagegen A Elementarereignisse der Menge E, die nicht zu B gehören, dann gehört A auch nicht zum Körper Z'. Ein gewisser

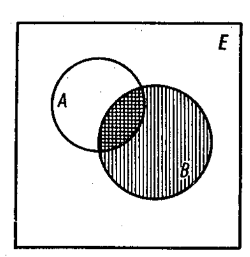


Abb. 1.5.1

Teil der Menge A kann aber ein zufälliges Ereignis in Z' sein, nämlich dann, wenn A und B gemeinsame Elemente in E besitzen, d. h., wenn das Produkt AB nicht die leere Menge ist.

Es bezeichne B jetzt ein bestimmtes Element des Mengenkörpers Z, wobei $P(B) > 0$, während A alle möglichen Elemente von Z durchlaufen soll. Alle Elemente von Z' sind dann Produkte der Form AB. Um zu betonen, daß wir das Produkt AB als Element von Z' (und nicht als Element von Z) betrachten, bezeichnen wir es mit dem Symbol $A \mid B$ und lesen: *A unter der Bedingung B* oder *A, wenn B eingetreten ist.*

Enthält A das Ereignis B, dann ist $A \mid B$ ein sicheres Ereignis (im Körper Z').

Wir wollen das Ereignis $A \mid B$ an Abb. 1.5.1 erläutern. Hier bedeuten das Quadrat E die Menge der Elementarereignisse und die Kreise A und B gewisse zufällige Ereignisse. Der schraffierte Bereich bedeutet das Ereignis B, während der karierte Bereich das zufällige Ereignis $A \mid B$ bedeutet (das Ereignis A, wenn das Ereignis B eingetroffen ist).

Die Wahrscheinlichkeit des Ereignisses $A \mid B$ werden wir im Körper Z' mit $P(A \mid B)$ bezeichnen und lesen: *die bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B* oder *wenn B eingetroffen ist.*

Diese Wahrscheinlichkeit kann man, wie gleich gezeigt wird, durch den Wahrscheinlichkeitsbegriff des Körpers Z bestimmen; man braucht also die Existenz der Wahrscheinlichkeit $P(A|B)$ und deren Eigenschaften nicht besonders zu fordern.

C. Um das Verständnis der Definition von $P(A|B)$ zu erleichtern, stellen wir folgende Überlegungen an:

Wir haben n Versuche unter gleichen Bedingungen ausgeführt und m-mal das Ereignis B erhalten. Außerdem haben wir noch bei k ($k \leqq m$) der m Versuche, bei denen das Ereignis B eintrat, als Ergebnis das zufällige Ereignis A erhalten. Die Häufigkeit des Produktes AB beträgt $\frac{k}{n}$, die des Ereignisses B beträgt $\frac{m}{n}$, während die Häufigkeit von A unter der Bedingung, daß B eingetroffen ist, $\frac{k}{m}$ beträgt.

Die Tatsache, daß die Gleichung

$$\frac{k}{m} = \frac{k}{n} : \frac{m}{n} \tag{1.5.1}$$

gilt, legt, wenn wir die Eigenschaft (1.5.1) der Häufigkeiten auf die Wahrscheinlichkeiten dieser Ereignisse übertragen, die folgende Definition nahe:

Definition 1.5.1. Die Wahrscheinlichkeit des Ereignisses B sei positiv. Die *bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B* ist gleich dem Quotienten aus der Wahrscheinlichkeit des Produktes AB und der Wahrscheinlichkeit des Ereignisses B.

Somit ist

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad \text{wenn} \quad P(B) > 0. \tag{1.5.2}$$

Analog ist

$$P(B|A) = \frac{P(AB)}{P(A)}, \quad \text{wenn} \quad P(A) > 0. \tag{1.5.3}$$

Aus den Formeln (1.5.2) und (1.5.3) erhalten wir

$$P(AB) = P(B)\,P(A|B) = P(A)\,P(B|A). \tag{1.5.4}$$

Diese Formel lesen wir: *Die Wahrscheinlichkeit des Produktes AB zweier Ereignisse ist gleich dem Produkt der Wahrscheinlichkeit des Ereignisses B mit der bedingten Wahrscheinlichkeit des Ereignisses A unter der Bedingung B* — oder — *gleich dem Produkt der Wahrscheinlichkeit von A mit der bedingten Wahrscheinlichkeit von B unter der Bedingung A.*

Mit A_1, A_2, A_3 seien drei zufällige Ereignisse bezeichnet, die derselben Menge Z der zufälligen Ereignisse angehören. Wir betrachten den Ausdruck $P(A_3 | A_1 A_2)$, die Wahrscheinlichkeit des Ereignisses A_3, wenn das Produkt $A_1 A_2$ eingetroffen ist. Nach (1.5.2) ist diese Wahrscheinlichkeit gleich

$$P(A_3 | A_1 A_2) = \frac{P(A_1 A_2 A_3)}{P(A_1 A_2)}, \tag{1.5.5}$$

wobei selbstverständlich $P(A_1 A_2) > 0$ vorausgesetzt ist.

Aus (1.5.5) und (1.5.3) erhalten wir für die Wahrscheinlichkeit des Produkts dreier Ereignisse die Beziehungen

$$P(A_1 A_2 A_3) = P(A_1 A_2)\, P(A_3 | A_1 A_2) = P(A_1)\, P(A_2 | A_1)\, P(A_3 | A_1 A_2). \tag{1.5.6}$$

Diese Formel lesen wir: *Die Wahrscheinlichkeit des Produkts dreier Ereignisse ist gleich der Wahrscheinlichkeit des ersten Ereignisses, multipliziert mit der bedingten Wahrscheinlichkeit des zweiten Ereignisses, wenn das erste eingetroffen ist, und der bedingten Wahrscheinlichkeit des dritten Ereignisses, wenn das Produkt der ersten beiden Ereignisse eingetroffen ist.*

Es seien $A_1, A_2, \ldots, A_n$ zufällige Ereignisse. Wir wollen die bedingte Wahrscheinlichkeit $P(A_{k_1} A_{k_2} \cdots A_{k_r} | A_{k_{r+1}} \cdots A_{k_n})$ des Produkts einer Untergruppe von r Ereignissen $(1 \leqq r \leqq n - 1)$ unter der Bedingung betrachten, daß das Produkt der übrigen $n - r$ Ereignisse eingetroffen ist. Eine ähnliche Überlegung wie vorher führt zur Formel

$$P(A_1 A_2 \cdots A_n) = P(A_1)\, P(A_2 | A_1)\, P(A_3 | A_1 A_2) \cdots P(A_n | A_1 \cdots A_{n-1}). \tag{1.5.7}$$

D. Wir wollen zeigen, daß die bedingte Wahrscheinlichkeit die Axiome I bis III erfüllt.

Die Ungleichung

$$P(AB) \leqq P(B) \tag{1.5.8}$$

folgt sofort aus Satz 1.3.6, da offensichtlich $AB \subset B$ ist.

Wegen $P(AB) \geqq 0$ und $P(B) > 0$ erhalten wir aus (1.5.8) die Formel

$$0 \leqq P(A | B) \leqq 1.$$

Also erfüllt die bedingte Wahrscheinlichkeit $P(A | B)$ das Axiom I.

Es sei $A | B$ das sichere Ereignis (im Körper Z'), also $AB = B$. In diesem Fall ist

$$P(AB) = P(B),$$

folglich

$$P(A \mid B) = 1.$$

Also ist auch das Axiom II erfüllt.

Wir betrachten die Summe $\sum_i (A_i \mid B)$ einander paarweise ausschließender Ereignisse. Wir können dafür

$$\sum_i (A_i \mid B) = \left(\sum_i A_i\right) \mid B,$$

also

$$P\left(\sum_i (A_i \mid B)\right) = P\left(\left(\sum_i A_i\right) \mid B\right)$$

schreiben. Aus Formel (1.5.2) sowie aus Axiom III folgt

$$P\left(\sum_i (A_i \mid B)\right) = \frac{P\left(\left(\sum_i A_i\right) B\right)}{P(B)} = \frac{P\left(\sum_i A_i B\right)}{P(B)} = \sum_i \frac{P(A_i B)}{P(B)} = \sum_i P(A_i \mid B).$$

Damit ist die vollständige Additivität der bedingten Wahrscheinlichkeiten, d. h. die Gültigkeit des Axioms III gezeigt.

Da die bedingte Wahrscheinlichkeit alle Axiome erfüllt, gelten auch alle Sätze, die man aus diesen Axiomen herleiten kann, für bedingte Wahrscheinlichkeiten.

1.6. Der Satz von Bayes

A. Bevor wir allgemeingültige Überlegungen anstellen, wollen wir ein Beispiel betrachten.

Beispiel 1.6.1. Wir haben zwei Urnen vor uns. In der ersten befinden sich drei weiße und zwei schwarze Kugeln und in der zweiten eine weiße und vier schwarze. Aus einer willkürlich gewählten Urne greifen wir blindlings eine Kugel heraus. Wie groß ist die Wahrscheinlichkeit, eine weiße Kugel herauszugreifen, wenn die Wahrscheinlichkeit, irgendeine der beiden Urnen zu wählen, je 0,5 beträgt?

Wir bezeichnen mit A_1 bzw. A_2 die entsprechenden Ereignisse, die darin bestehen, daß wir irgendeine Kugel aus der ersten bzw. aus der zweiten Urne herausgreifen, und mit B das Ereignis, daß wir eine weiße Kugel herausgreifen. Das Ereignis B kann entweder gemeinsam mit dem Ereignis A_1 oder mit dem Ereignis A_2 eintreten. Somit haben wir

$$B = A_1 B + A_2 B$$

und, da sich die Ereignisse $A_1 B$ und $A_2 B$ gegenseitig ausschließen,

$$P(B) = P(A_1 B) + P(A_2 B).$$

Wenn wir die Formel (1.5.4) anwenden, erhalten wir

$$P(B) = P(A_1)\, P(B \mid A_1) + P(A_2)\, P(B \mid A_2). \tag{1.6.1}$$

Für unser Beispiel ist $P(A_1) = P(A_2) = 0{,}5$, $P(B|A_1) = 0{,}6$ und $P(B|A_2) = 0{,}2$. Setzen wir diese Werte in die Formel (1.6.1) ein, so erhalten wir $P(B) = 0{,}4$.

Die im letzten Beispiel erhaltene Formel ist ein Sonderfall des Satzes über die *totale Wahrscheinlichkeit*, den wir jetzt angeben wollen:

Satz 1.6.1. *Wenn die zufälligen Ereignisse $A_1, A_2, \ldots$ einander paarweise ausschließen und die Menge E der Elementarereignisse ausschöpfen, wobei $P(A_i) > 0$ für $i = 1, 2, \ldots$, dann gilt für ein beliebiges zufälliges Ereignis B die Gleichung*

$$P(B) = P(A_1)\,P(B|A_1) + P(A_2)\,P(B|A_2) + \cdots. \tag{1.6.2}$$

In der Tat folgt aus den Voraussetzungen, daß das Ereignis B nur gemeinsam mit genau einem der Ereignisse A_i eintreffen kann. Es ist also

$$B = A_1B + A_2B + \cdots$$

sowie

$$P(B) = P(A_1B) + P(A_2B) + \cdots. \tag{1.6.3}$$

Laut (1.5.4) erhalten wir für jedes i die Gleichung

$$P(A_iB) = P(A_i)P(B|A_i). \tag{1.6.4}$$

Wenn wir für $P(A_iB)$ die Ausdrücke (1.6.4) in die Formel (1.6.3) einsetzen, erhalten wir (1.6.2).

B. Im folgenden mögen die Ereignisse A_i und B die in der Voraussetzung zu Satz 1.6.1 aufgeführten Eigenschaften haben. Wir stellen uns vor, daß das Ereignis B eingetreten ist. Wie groß ist dann die Wahrscheinlichkeit des Ereignisses A_i? Eine Antwort auf diese Frage gibt der folgende Satz von Bayes:

Satz 1.6.2. *Erfüllen $A_1, A_2, \ldots$ und B die Voraussetzungen des Satzes über die totale Wahrscheinlichkeit und ist $P(B) > 0$, so gilt für $i = 1, 2, \ldots$ die Formel*

$$P(A_i|B) = \frac{P(A_i)\,P(B|A_i)}{P(A_1)\,P(B|A_1) + P(A_2)\,P(B|A_2) + \cdots}. \tag{1.6.5}$$

Setzen wir nämlich in der Formel (1.5.4) A_i statt A, so erhalten wir

$$P(A_i|B) = \frac{P(A_i)\,P(B|A_i)}{P(B)}.$$

Wenn wir noch im Nenner den Ausdruck $P(B)$ durch die rechte Seite der Formel (1.6.2) ersetzen, erhalten wir (1.6.5).

Die Formel (1.6.5) ist nach Bayes benannt. Man bezeichnet sie auch als Formel für die *Wahrscheinlichkeit a posteriori*. Dieser Name erklärt sich daraus, daß

diese Formel die Wahrscheinlichkeit der Realisierung des Ereignisses A_i nach dem Eintreffen von B angibt. Im Gegensatz zur Wahrscheinlichkeit $P(A_i|B)$ nennen wir den in dieser Formel auftretenden Ausdruck $P(A_i)$ *Wahrscheinlichkeit a priori*.

In den Anwendungen spielt die Bayessche Formel eine große Rolle.

Beispiel 1.6.2. Die Geschütze Nr. 1 und Nr. 2 feuern auf dasselbe Ziel. Man hat festgestellt, daß das Geschütz Nr. 1 durchschnittlich neun Schüsse abgibt in der Zeit, in der das Geschütz Nr. 2 zehn Geschosse abfeuert. Die Treffsicherheit der beiden Geschütze sei nicht dieselbe; vom Geschütz Nr. 1 treffen nämlich durchschnittlich 8 von 10 Geschossen das Ziel, während vom Geschütz Nr. 2 nur 7 von 10 das Ziel treffen.

Während des Beschusses wurde nun das Ziel von einem Geschoß getroffen, man weiß aber nicht, aus welchem Geschütz dieses Geschoß stammt. Wie groß ist die Wahrscheinlichkeit dafür, daß dieses Geschoß vom Geschütz Nr. 2 abgeschossen wurde?

Wir bezeichnen mit A_1 bzw. A_2 die Ereignisse, die darin bestehen, daß das Geschütz Nr. 1 bzw. Nr. 2 ein Geschoß abschießt. Das Verhältnis der durchschnittlichen Geschoßanzahl des Geschützes Nr. 1 pro Zeiteinheit zur durchschnittlichen Geschoßanzahl des Geschützes Nr. 2 pro Zeiteinheit beträgt $9/10 = 0{,}9$. Wir nehmen deshalb an, daß $P(A_1) = 0{,}9P(A_2)$ ist.[1] Weiter bezeichne B das Ereignis, das darin besteht, daß irgendein Geschoß das Ziel trifft. Auf Grund der Angaben, die wir über die Treffsicherheit der Geschütze besitzen, setzen wir $P(B|A_1) = 0{,}8$ und $P(B|A_2) = 0{,}7$.

Die Bayessche Formel ergibt

$$P(A_2|B) = \frac{P(A_2)P(B|A_2)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$

$$= \frac{0{,}7P(A_2)}{0{,}9P(A_2)\cdot 0{,}8 + 0{,}7P(A_2)} = 0{,}493.$$

1.7. Unabhängige Ereignisse

A. Die bedingte Wahrscheinlichkeit $P(A|B)$ ist im allgemeinen verschieden von $P(A)$. Besonders wichtig ist nun der Fall, daß beide gleich sind:

$$P(A|B) = P(A). \tag{1.7.1}$$

In diesem Fall hat die Tatsache, daß das Ereignis B eingetroffen ist, keinerlei Einfluß auf die Wahrscheinlichkeit des Ereignisses A, oder anders gesagt: Die Wahrscheinlichkeit des Ereignisses A hängt nicht davon ab, ob das Ereignis B eingetroffen ist oder nicht.

Aus (1.5.4) erhalten wir, wenn die Beziehung (1.7.1) erfüllt ist, die Gleichung

$$P(AB) = P(A)\,P(B). \tag{1.7.2}$$

[1]) Im zweiten Teil dieses Buches geben wir Methoden an, die es erlauben, die Berechtigung derartiger Hypothesen nachzuprüfen.

Die Formel (1.7.2) gilt auch dann, wenn die Gleichung

$$P(B|A) = P(B) \tag{1.7.3}$$

besteht.

Wir leiteten (1.7.2) aus der Formel (1.5.4) her, wobei für die letzte die Annahmen $P(A) > 0$ und $P(B) > 0$ zu machen waren; die Formel (1.7.2) ist aber auch in dem Fall sinnvoll, daß eine dieser Wahrscheinlichkeiten gleich Null ist.

Wir definieren jetzt die Unabhängigkeit zweier zufälliger Ereignisse.

Definition 1.7.1. Die Ereignisse A und B sind *voneinander unabhängig*, wenn für ihre Wahrscheinlichkeiten die Formel (1.7.2) gilt, d. h., wenn die Wahrscheinlichkeit des Produkts AB gleich dem Produkt der Wahrscheinlichkeiten der Ereignisse A und B ist.

Wie aus dieser Definition folgt, ist der Begriff der Unabhängigkeit zweier zufälliger Ereignisse symmetrisch in bezug auf diese Ereignisse.

Wie wir schon bemerkten, können wir aus jeder der Formeln (1.7.1) und (1.7.3) die Formel (1.7.2) erhalten. Aus den Formeln (1.5.4) und (1.7.2) folgen unter der Annahme, daß $P(A) > 0$ und $P(B) > 0$, ebenfalls die Formeln (1.7.1) und (1.7.3). Daraus schließen wir: Jede der beiden letzten Formeln ist eine notwendige und hinreichende Bedingung für die Unabhängigkeit von zwei zufälligen Ereignissen, deren Wahrscheinlichkeiten positiv sind.

B. Man kann die Unabhängigkeitsdefinition auf endlich oder abzählbar viele zufällige Ereignisse ausdehnen.

Definition 1.7.2. Die zufälligen Ereignisse $A_1, A_2, \ldots, A_n$ sind *voneinander unabhängig*, wenn für beliebige ganze Indizes $k_1, k_2, \ldots, k_s$, die die Bedingungen

$$1 \leqq k_1 < k_2 < \cdots < k_s \leqq n$$

erfüllen, die Gleichung

$$P(A_{k_1} A_{k_2} \cdots A_{k_s}) = P(A_{k_1}) P(A_{k_2}) \cdots P(A_{k_s})$$

gilt, d. h., wenn die Wahrscheinlichkeit des Produkts $A_{k_1} A_{k_2} \cdots A_{k_s}$ jeder Kombination von Ereignissen gleich dem Produkt der Wahrscheinlichkeiten dieser Ereignisse ist.

Es ist möglich, daß die Ereignisse $A_1, A_2, \ldots, A_n$ zwar paarweise unabhängig sind (d. h., daß je zwei der Ereignisse voneinander unabhängig sind) oder zu dreien unabhängig sind (d. h., daß je drei Ereignisse voneinander unabhängig sind), die Ereignisse $A_1, A_2, \ldots, A_n$ insgesamt aber nicht voneinander unabhängig sind.

Ein Beispiel, das von S. Bernstein angegeben wurde, soll das erläutern:

Beispiel 1.7.1. In einer Urne befinden sich 4 Papierstreifen gleicher Größe. Jeder Papierstreifen ist mit einer der vier folgenden Aufschriften versehen: 110, 101, 011, 000, wobei nicht zwei Streifen die gleiche Aufschrift tragen. Das Ereignis A_1 bestehe im Herausgreifen eines

Streifens mit einer Aufschrift, die an erster Stelle eine Eins hat, das Ereignis A_2 im Ziehen eines Streifens mit einer Eins an zweiter Stelle, A_3 im Ziehen eines Streifens mit einer Eins an dritter Stelle. Da für jede dieser Kategorien zwei Streifen vorhanden sind und da die Gesamtzahl der Streifen 4 beträgt, erhalten wir unter der Voraussetzung, daß die Wahrscheinlichkeit, gezogen zu werden, für alle Streifen dieselbe ist,

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}.$$

Bezeichnen wir mit A das Produkt $A_1 A_2 A_3$, so ist $P(A) = 0$; denn das Ereignis A ist unmöglich (auf keinem einzigen Papierstreifen befinden sich drei Einsen!). Wegen

$$P(A_1)P(A_2)P(A_3) = \frac{1}{8} \neq 0 = P(A)$$

sind die Ereignisse A_1, A_2, A_3 nicht voneinander unabhängig. Aber wir werden zeigen, daß diese Ereignisse paarweise voneinander unabhängig sind. Für das Paar A_1, A_2 gilt nämlich

$$P(A_2 | A_1) = \frac{1}{2} = P(A_2);$$

denn es gibt nur zwei Streifen, die an erster Stelle eine Eins haben, und nur einer von ihnen hat an zweiter Stelle eine Eins. Ähnlich kann man die Unabhängigkeit der übrigen Paare nachweisen.

Die Unabhängigkeit abzählbar vieler zufälliger Ereignisse erklären wir folgendermaßen:

Definition 1.7.3. Die zufälligen Ereignisse $A_1, A_2, A_3, \ldots$ sind *voneinander unabhängig*, wenn für jedes $n = 2, 3, 4, \ldots$ die Ereignisse $A_1, \ldots, A_n$ voneinander unabhängig sind.

Wie aus den Definitionen 1.7.3 und 1.7.2 folgt, sind, falls die zufälligen Ereignisse $A_1, A_2, A_3, \ldots$ voneinander unabhängig sind, auch für jedes $n = 2, 3, 4, \ldots$ und für beliebige Indizes $k_1, k_2, \ldots, k_n$ die Ereignisse $A_{k_1}, \ldots, A_{k_n}$ voneinander unabhängig.

Um zu unterstreichen, daß die betrachteten Ereignisse nicht nur paarweise oder zu dreien usw. unabhängig sind, sondern im Sinne der Definition 1.7.2 oder 1.7.3, wird in der Literatur zur Wahrscheinlichkeitsrechnung die Bezeichnung *unabhängig en bloc* oder *in der Gesamtheit unabhängig* oder *insgesamt unabhängig* gebraucht. Diese Ausdrücke wollen wir vermeiden und kurz *unabhängig* statt *unabhängig en bloc* sagen.

1.8. Aufgaben und Ergänzungen

1. Man zeige, daß die Operationen der Addition und Multiplikation zufälliger Ereignisse kommutativ sind und daß sie dem Assoziativgesetz sowie dem Distributivgesetz genügen:

$$A_1 + A_2 = A_2 + A_1,$$

$$A_1 A_2 = A_2 A_1,$$

$$A_1 + (A_2 + A_3) = (A_1 + A_2) + A_3,$$
$$A_1(A_2A_3) = (A_1A_2)A_3,$$
$$A_1(A_2 + A_3) = A_1A_2 + A_1A_3.$$

2. Man beweise die Gleichungen

$$A_1A_2 = A_1 - (A_1 - A_2),$$
$$A_1 + (A_2 - A_1) = A_1 + A_2,$$
$$A_1 - A_2 = A_1 - A_1A_2,$$
$$A_1(A_2 - A_3) = A_1A_2 - A_3.$$

3. Man zeige, daß $A_1 + A_2 = A_2$ und $A_1A_2 = A_1$ gilt, wenn $A_1 \subset A_2$ ist.

4. a) Man zeige die Gültigkeit der *de Morganschen Formeln*

$$\overline{A_1 + A_2} = \overline{A_1}\,\overline{A_2},$$
$$\overline{A_1A_2} = \overline{A_1} + \overline{A_2}.$$

b) Man verallgemeinere diese Formeln für den Fall von n $(n > 2)$ zufälligen Ereignissen.

5. a) Man zeige, daß für $n = 2, 3, \ldots$ die Gleichung

$$A_1 + \cdots + A_n = A_1 + (A_2 - A_1A_2) + \cdots + (A_n - A_1A_n - \cdots - A_{n-1}A_n)$$

gilt. Wir weisen darauf hin, daß sich die Summanden auf der rechten Seite paarweise aufheben.

b) Man zeige die Gültigkeit der Beziehung

$$\bigcup_{i=1}^{\infty} A_i = A_1 + (A_2 - A_1A_2) + (A_3 - A_1A_3 - A_2A_3) + \cdots.$$

6. Man beweise, daß die Eigenschaften 1.2.2 und 1.2.5 aus den Eigenschaften 1.2.1, 1.2.3 und 1.2.4 folgen.

7. Es sei $\{A_n\}$ $(n = 1, 2, \ldots)$ eine Folge zufälliger Ereignisse. Das zufällige Ereignis A^*, das sämtliche Elementarereignisse enthält, die zu unendlich vielen A_n gehören, heißt *oberer Grenzwert* der Folge $\{A_n\}$:

$$A^* = \limsup_n A_n.$$

Das zufällige Ereignis A_*, das sämtliche Elementarereignisse enthält, die zu sämtlichen A_n mit Ausnahme von endlich vielen gehören, heißt *unterer Grenzwert* der Folge $\{A_n\}$:

$$A_* = \liminf_n A_n.$$

a) Man zeige, daß die Beziehungen

$$A^* = \prod_{n=1}^{\infty} \sum_{k=n}^{\infty} A_k, \quad A_* = \sum_{n=1}^{\infty} \prod_{k=n}^{\infty} A_k$$

gelten.

b) Man beweise die Beziehung

$$A_* \subset A^*.$$

8. Mit den Bezeichnungen der vorhergehenden Aufgabe nennt man $A = A^* = A_*$, falls $A^* = A_*$ gilt, den *Grenzwert* der Folge $\{A_n\}$ und schreibt

$$A = \lim_{n\to\infty} A_n.$$

a) Man zeige, daß

$$A = A_* = A^* = \sum_{n=1}^{\infty} A_n$$

gilt, wenn $\{A_n\}$ eine zunehmende Folge ist.

b) Man zeige, daß

$$A = A_* = A^* = \prod_{n=1}^{\infty} A_n$$

gilt, wenn $\{A_n\}$ eine abnehmende Folge ist.

9. a) Man zeige, daß für eine beliebige Folge $\{A_n\}$ von zufälligen Ereignissen die folgenden Beziehungen gelten:

$$P\left(\limsup_{n\to\infty} A_n\right) \geqq \limsup_{n\to\infty} P(A_n),$$

$$P\left(\liminf_{n\to\infty} A_n\right) \leqq \liminf_{n\to\infty} P(A_n).$$

Hinweis. Man verwende die Sätze 1.3.4 und 1.3.5.

b) Man beweise, daß

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

gilt, wenn der Grenzwert $\lim\limits_{n\to\infty} A_n$ existiert.

10. Unter Verwendung der Aufgaben 5a und 5b beweise man die Ungleichungen

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqq \sum_{i=1}^{n} P(A_i), \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) \leqq \sum_{i=1}^{\infty} P(A_i).$$

In den nachstehenden kombinatorischen Aufgaben 11 bis 18 ist anzunehmen, daß sämtliche Systeme gleichwahrscheinlich sind.

11. Ein Kartenspiel möge 52 Karten enthalten. Ein Spieler G erhielt 13 Karten. Man berechne die Wahrscheinlichkeit dafür, daß G folgende Karten erhält:

a) genau drei Asse,
b) wenigstens drei Asse,
c) irgendwelche drei gleiche Bilder,
d) irgendwelche drei gleiche Karten aus den fünf höchsten Nennwerten,
e) irgendwelche drei gleiche Karten aus den acht niedrigsten Nennwerten,
f) irgendwelche drei Karten mit dem gleichen Nennwert,
g) drei aufeinanderfolgende Karten der Farbe Pique,
h) wenigstens drei aufeinanderfolgende Karten der Farbe Pique,
i) drei aufeinanderfolgende Karten derselben (beliebigen Farbe),
k) wenigstens drei aufeinanderfolgende Karten derselben (beliebigen) Farbe.

12. Drei Würfel werden einmal geworfen. Man berechne die Wahrscheinlichkeit dafür, daß man die folgenden Augenkombinationen erhält:

    a) auf einem Würfel zwei Augen,
    b) auf wenigstens einem Würfel drei Augen,
    c) eine gerade Augensumme,
    d) eine durch drei teilbare Augensumme,
    e) eine Augensumme, die größer ist als sieben,
    f) eine Augensumme, die kleiner ist als zwölf,
    g) eine Augensumme, die Primzahl ist.

13. (*Die Aufgabe des Chevalier de Méré.*) Man stelle fest, welches der folgenden Ereignisse die größere Wahrscheinlichkeit hat:

    a) bei einem gleichzeitigen Wurf mit vier Würfeln auf wenigstens einem Würfel die Augenzahl 1 zu erhalten;
    b) bei 24 Würfen mit gleichzeitig zwei Würfeln wenigstens einmal zwei Augen zu erhalten.

14. (*Das Banachsche Problem.*) Ein Mathematiker trägt zwei Streichholzschachteln bei sich, von denen jede ursprünglich n Streichhölzer enthielt. Jedesmal, wenn er sich eine Zigarette anzündet, wählt er eine Streichholzschachtel zufällig aus. Man berechne die Wahrscheinlichkeit dafür, daß bei eventueller Auswahl einer bereits leeren Schachtel in der anderen Schachtel noch r Streichhölzer vorhanden sind, $r = 0, 1, \ldots, n$.

15. Eine Urne enthält m weiße und $n - m$ schwarze Kugeln. Zwei Spieler greifen nacheinander beliebig eine Kugel aus der Urne heraus und tun diese vor der Entnahme der nächsten Kugel in die Urne zurück. Der Spieler, dem es zuerst gelingt, eine weiße Kugel zu ziehen, hat das Spiel gewonnen. Man berechne die Wahrscheinlichkeit dafür, daß der beginnende Spieler gewinnt.

16. In einer Urne mögen sich n Kugeln befinden. Wir entnehmen aus ihr eine beliebige Anzahl. Man zeige, daß die Wahrscheinlichkeit, eine gerade Anzahl von Kugeln zu ziehen, sich zu

$$\frac{2^{n-1} - 1}{2^n - 1}$$

ergibt.

17. Eine Urne enthält n weiße und n schwarze Kugeln. Wir entnehmen der Urne auf beliebige Art und Weise eine gerade Anzahl von Kugeln. Man zeige, daß die Wahrscheinlichkeit, eine gleiche Anzahl weißer und schwarzer Kugeln zu ziehen,

$$\frac{\frac{(2n)!}{(n!)^2} - 1}{2^{2n-1} - 1}$$

beträgt.

18. Auf 28 Karten ist je ein Buchstabe eingetragen. Die dabei auftretenden Buchstaben und deren jeweilige Vielfachheit sind in der nachstehenden Tabelle angegeben:

Buchstabe	a	c	e	h	i	j	l	m	n	o	s	t	y	ż
Vielfachheit	3	1	3	1	1	2	2	1	2	2	2	4	2	2

Die Karten werden dann in einer beliebigen Reihenfolge angeordnet. Wie groß ist die Wahrscheinlichkeit dafür, daß sich hierbei der Satz: „sto lat, sto lat niech żyje, żyje nam" ergibt?[1])

19. GOETHE schenkte einst dem bekannten Chemiker RUNGE, der bei ihm zu Besuch weilte, eine Büchse mit Kaffeebohnen, ein Geschenk, das zu jener Zeit äußerst wertvoll war. RUNGE verwandte seinen Inhalt für wissenschaftliche Zwecke, und es gelang ihm als erstem, reines Koffein darzustellen. Läßt sich die Wahrscheinlichkeit dieses Ereignisses berechnen und, wenn ja, ist dieses Ergebnis eindeutig? Wovon hängt die präzise Formulierung des zufälligen Ereignisses ab, dessen Wahrscheinlichkeit wir suchen?

20. (*Das Bertrandsche Paradoxon.*) Einem Kreis sei ein gleichseitiges Dreieck mit der Seitenlänge a einbeschrieben. Dann wird im Kreis eine beliebige Sehne gezogen. Das Ereignis A besteht darin, daß die Länge l der Sehne der Bedingung $l > a$ genügt. Man ermittle die Bedingungen, unter denen a) $P(a) = 0{,}5$, b) $P(a) = \frac{1}{3}$ und c) $P(a) = \frac{1}{4}$ ist. Sind diese Ergebnisse als widersprüchlich aufzufassen?

21. (*Die Buffonsche Aufgabe.*) Eine Nadel von der Länge $2\,l$ wird in beliebiger Weise auf eine waagerechte Ebene geworfen, in der parallele Geraden im Abstand $2a$ $(a > l)$ gezogen sind. Wie groß ist die Wahrscheinlichkeit dafür, daß die Nadel eine der Geraden schneidet?

22. Die Wahrscheinlichkeit dafür, daß unter Zwillingen beide Kinder Knaben sind, ist 0,32, während die Wahrscheinlichkeit dafür, daß beide Mädchen sind, 0,28 beträgt. Man bestimme die bedingte Wahrscheinlichkeit dafür, daß

    a) das zweite Kind unter Zwillingen ein Knabe ist, wenn bereits das erste ein Knabe ist;

    b) das zweite Zwillingskind ein Mädchen ist, wenn bereits das erste ein Mädchen ist.

    Hinweis. Man verwende das Beispiel 1.1.3.

23. a) Wie oft muß ein Würfel geworfen werden, damit die Wahrscheinlichkeit für das wenigstens einmalige Auftreten der Augenzahl 6 nicht kleiner als $\frac{3}{4}$ ist?

    b) Die Ereignisse $A_1, A_2, \ldots$ seien unabhängig, wobei $P(A_j) = p$ $(j = 1, 2, \ldots)$ gilt. Man bestimme den Wert n, für den die Beziehung

$$P\left(\bigcup_{k=1}^{n} A_k\right) \geqq p_0$$

    erfüllt ist, wobei p_0 eine beliebig vorgegebene Zahl bedeutet.

24. a) Die Ereignisse $A_1, A_2, \ldots, A_n$ seien unabhängig, und es sei hierbei $P(A_k) = p_k$. Wie groß ist die Wahrscheinlichkeit dafür, daß keines der Ereignisse A_k eintritt?

    b) Man lasse die Unabhängigkeitsvoraussetzung fallen und beantworte dann die Frage.

25. Man zeige, daß bei unabhängigen zufälligen Ereignissen A und B auch $\bar{A}$ und $\bar{B}$ unabhängig sind.

---

[1]) Es handelt sich um den Anfang eines polnischen Liedes, das unserem „Hoch soll er leben" entspricht und etwa lautet: „Hundert Jahre, hundert Jahre alt möge er werden."

2. ZUFALLSVARIABLE

2.1. Der Begriff einer Zufallsvariablen

Jedem Elementarereignis, das zu einer gewissen Menge E gehört, können wir irgendeine Zahl zuordnen. So ordnen wir z. B. beim Werfen einer Münze dem Auftreten des Adlers die Eins und dem Auftreten der Zahl die Null zu. Die Wahrscheinlichkeit dafür, daß wir als Versuchsergebnis die Eins erhalten, ist gleich der Wahrscheinlichkeit, daß die Münze beim Auffallen den Adler zeigt; ebenso ist die Wahrscheinlichkeit, die Null zu erhalten, gleich der Wahrscheinlichkeit dafür, daß die Münze beim Auffallen die Zahl zeigt.

Ebenso kann man beim Würfeln den einzelnen Elementarereignissen, die hier im Auftreten einer der Augenzahlen 1, 2, 3, 4, 5, 6 bestehen, eben diese Zahlen zuordnen. Allgemein sei durch e ein Elementarereignis bezeichnet, das zur Menge E der Elementarereignisse gehört. Auf der Menge E definieren wir eine eindeutige reelle Funktion derart, daß — wir drücken uns vorläufig ganz unexakt aus — die Wahrscheinlichkeit, mit der diese Funktion gewisse Werte annimmt, bestimmt ist. Um die Bedingungen, denen diese Funktion genügen soll, genauer formulieren zu können, führen wir erst den Begriff des Urbildes ein.

Definition 2.1.1. Es sei $X(e)$ eine eindeutige reelle Funktion, die auf der Menge E der Elementarereignisse definiert ist. Als *Urbild* einer gegebenen Menge S von reellen Zahlen bezeichnen wir die Menge A aller derjenigen Elementarereignisse e, denen die Funktion $X(e)$ Werte aus der Menge S zuordnet.

Wir schreiben hierfür

$$A = X^{-1}(S).$$

Man sieht leicht, daß das Urbild der Menge R aller reellen Zahlen die Menge E ist: $E = X^{-1}(R)$.

Ist S eine gewisse Menge reeller Zahlen auf der x-Achse, so bedeute das Symbol $P^{(x)}(S)$ die Wahrscheinlichkeit dafür, daß die Funktion $X(e)$ einen Wert aus S annimmt:

$$P^{(x)}(S) = P^{(x)}(X(e) \in S).$$

Definition 2.1.2. Wir nennen eine eindeutige reelle Funktion $X(e)$, die auf der Menge E aller Elementarereignisse e definiert ist, eine *Zufallsvariable*[1]),

[1]) Dem Begriff einer Zufallsvariablen entspricht in der Theorie der reellen Funktionen der Begriff einer (hinsichtlich des betrachteten Mengenkörpers) meßbaren Funktion.

wenn das Urbild A eines jeden reellen Zahlenintervalls I der Form $(-\infty, x)$ ein zufälliges Ereignis ist.

Die Wahrscheinlichkeit $P^{(x)}(I)$ dafür, daß die Zufallsvariable $X(e)$ einen Wert aus dem Intervall I annimmt, setzen wir gleich der Wahrscheinlichkeit $P(A)$, wobei A das Urbild des Intervalls I ist.

Von jetzt an wollen wir gewöhnlich X statt $X(e)$ schreiben und uns merken, daß die Funktion X auf der Menge E der Elementarereignisse definiert ist.

Zufallsvariable bezeichnen wir gewöhnlich mit großen Buchstaben $X, Y, \ldots$, während wir die Werte, die sie annehmen, mit den entsprechenden kleinen Buchstaben $x, y, \ldots$ bezeichnen.

Aus der Definition 2.1.2 folgt insbesondere: Ist das zufällige Ereignis A das Urbild eines Punktes x, dann ist die Wahrscheinlichkeit dafür, daß die Zufallsvariable X den Wert x annimmt, gleich der Wahrscheinlichkeit des Ereignisses A:

$$P^{(x)}(X = x) = P(A).$$

Da ein jedes halboffene Intervall I der Form $[a, b)$ mit $a < b$ gleich der Differenz der Intervalle $(-\infty, b)$ und $(-\infty, a)$ ist, folgt aus der Definition 2.1.2 die Existenz der Wahrscheinlichkeit $P^{(x)}(I)$. Wie man zeigen kann, ist das Urbild der Vereinigung, der Differenz bzw. des Durchschnitts von beliebigen Intervallen gleich der Vereinigung, der Differenz bzw. dem Durchschnitt der Urbilder dieser Intervalle.

Jede Borelsche Menge von reellen Zahlen kann man als Ergebnis endlich oder abzählbar vieler Operationen wie Vereinigungsbildung, Differenzbildung oder Durchschnittsbildung, ausgeführt an rechtsseitig halboffenen Intervallen, auffassen (siehe auch Anhang, Beispiel A 2).

Aus den Eigenschaften der Menge Z der zufälligen Ereignisse sowie aus den Eigenschaften der Wahrscheinlichkeit, die in Kapitel 1 besprochen wurden, und endlich aus der Definition der Zufallsvariablen folgt: *Ist X eine Zufallsvariable, so ist für jede Borelsche Menge S auf der x-Achse die Wahrscheinlichkeit $P^{(x)}(S)$ dafür definiert, daß X einen Wert annimmt, der der Menge S angehört.* Mit anderen Worten: *Die Wahrscheinlichkeitsverteilung der Zufallsvariablen X ist bestimmt* (siehe Anhang, Satz über die Erweiterung). Statt *Wahrscheinlichkeitsverteilung der Zufallsvariablen X* wollen wir kürzer *Verteilung der Zufallsvariablen X* sagen.

Definition 2.1.3. Die Funktion $P^{(x)}(S)$, die gleich der Wahrscheinlichkeit dafür ist, daß die Zufallsvariable X einen zu S gehörigen Wert annimmt, wobei S eine beliebige Borelsche Menge der x-Achse ist, nennen wir *die Wahrscheinlichkeitsfunktion der Zufallsvariablen X.*

Wir schreiben

$$P^{(x)}(S) = P^{(x)}(X \in S).$$

In den Anwendungen werden wir es meistens mit einfachen Zahlenmengen zu tun haben, z. B. mit Intervallen oder endlichen Summen von Intervallen.

Aus der Definition der Zufallsvariablen ersieht man, daß sie eine *Funktion* und nicht eine Variable im gewöhnlichen Sinne der Analysis ist. Diese Bemerkung ist sehr wichtig und ständig zu beachten. Insbesondere muß man sehr vorsichtig sein, wenn man formale Eigenschaften von Operationen mit Variablen auf Operationen mit Zufallsvariablen übertragen will.

Hier in 2.1, in 2.5.A und im Anhang verwenden wir das Symbol $P^{(x)}$ zur Bezeichnung der Wahrscheinlichkeiten für Zufallsvariable und damit für Ereignisse, die Teilmengen einer neuen Menge $X(E)$ von Elementarereignissen sind, während das Symbol P für zufällige Ereignisse verwendet wurde, die Teilmengen der ursprünglichen Menge E waren. Im übrigen Teil des Buches ist eine derartige Unterscheidung nicht erforderlich, weshalb dort durchgängig das Symbol P benutzt wird.

2.2. Die Verteilungsfunktion

Es ist bequem, die Wahrscheinlichkeitsverteilung mit Hilfe einer sogenannten *Verteilungsfunktion* zu charakterisieren. Wir wollen diesen Begriff an einem Beispiel erläutern.

Beispiel 2.2.1. Wir betrachten Würfe mit einem Würfel. Den Elementarereignissen, die im Erscheinen der Zahlen 1, 2, 3, 4, 5, 6 bestehen, ordnen wir die gewürfelte Zahl zu. Die Zufallsvariable X kann hier die sechs Werte $x_i = i$ $(i = 1, \ldots, 6)$ mit derselben Wahrscheinlichkeit $P(X = x_i) = 1/6$ annehmen.

Die Wahrscheinlichkeit dafür, daß $X < 1$, ist natürlich gleich Null:

$$P(X < 1) = 0.$$

Wenn x eine Zahl ist, die der Ungleichung $1 < x \leqq 2$ genügt, so ist

$$P(X < x) = P(X = 1) = \frac{1}{6}.$$

Falls $2 < x \leqq 3$ ist, so sehen wir leicht ein, daß

$$P(X < x) = P(X = 1) + P(X = 2) = \frac{1}{3}.$$

Wenn wir schließlich $5 < x \leqq 6$ nehmen, erhalten wir

$$P(X < x) = \sum_{i=1}^{5} P(X = i) = \frac{5}{6}$$

und endlich für $x > 6$

$$P(X < x) = P(X \leqq 6) = \sum_{i=1}^{6} P(X = i) = 1.$$

Stellen wir bei diesem Beispiel die Werte $P(X < x)$ als Funktion der Variablen x graphisch dar, so erhalten wir eine Treppenlinie (Abb. 2.2.1).

Wie wir sehen, nimmt mit wachsendem x der Wert $P(X < x)$ nicht ab, und in den Punkten x_i $(i = 1, \ldots, 6)$ wächst dieser Wert um die Größe $P(X = x_i)$ an.

Definition 2.2.1. Die durch die Formel

$$F(x) = P(X < x) \tag{2.2.1}$$

definierte Funktion nennen wir die *Verteilungsfunktion der Zufallsvariablen* X.[1])

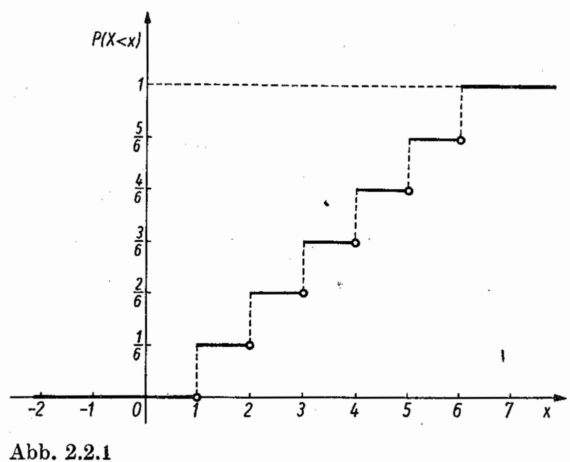


Abb. 2.2.1

Wie man sieht, bestehen die Gleichungen

$$F(-\infty) = 0, \qquad F(\infty) = 1. \tag{2.2.2}$$

Wir zeigen, daß die Verteilungsfunktion $F(x)$ nicht fallend ist. Es seien x_1 und x_2 Punkte auf der x-Achse, wobei $x_1 < x_2$. Da das Intervall $(-\infty, x_2)$ das Intervall $(-\infty, x_1)$ enthält, ist $P(X < x_2) \geqq P(X < x_1)$, also

$$F(x_2) \geqq F(x_1).$$

Wir zeigen jetzt, daß die Verteilungsfunktion wenigstens linksseitig stetig ist. Es sei $x_1 < x_2 < \cdots < x$ eine beliebige wachsende und gegen x konvergierende Zahlenfolge. Mit A_k bezeichnen wir das Ereignis, daß die Zufallsvariable X einen Wert annimmt, der zum halboffenen Intervall $[x_k, x)$ gehört. Ist $k_1 < k_2$, so folgt aus dem Eintreffen des Ereignisses A_{k_2} das Eintreffen von A_{k_1}. Es ist also $\{A_k\}$ eine absteigende Folge von Ereignissen. Der Grenzwert der Folge $\{x_k\}$, der

---

[1]) Manche Autoren definieren die Verteilungsfunktion durch die Formel

$$F(x) = P(X \leqq x).$$

Punkt x, gehört zu keinem der betrachteten Intervalle. Deshalb ist es unmöglich, daß die Zufallsvariable X einen zu allen Intervallen $[x_k, x)$ gehörigen Wert annimmt. Daraus folgt, daß das Produkt $A = \prod_{k=1}^{\infty} A_k$ ein unmögliches Ereignis ist, daß also $P(A) = 0$. Somit erhalten wir aus Satz 1.3.4

$$\lim_{k\to\infty} P(A_k) = \lim_{k\to\infty} P(x_k \leqq X < x) = \lim_{k\to\infty} [F(x) - F(x_k)]$$
$$= F(x) - \lim_{k\to\infty} F(x_k) = 0$$

oder

$$\lim_{k\to\infty} F(x_k) = F(x).$$

Da die Folge $\{x_k\}$ eine beliebige wachsende, von links gegen x konvergierende Folge war, folgt aus der letzten Gleichung die linksseitige Stetigkeit von $F(x)$.

Wie wir jetzt zeigen wollen, bestimmt jede reelle nichtfallende und wenigstens linksseitig stetige Funktion $F(x)$, die den Beziehungen (2.2.2) genügt, die Wahrscheinlichkeitsverteilung einer gewissen Zufallsvariablen. Dazu nehmen wir an, $F(x)$ erfülle die aufgezählten Bedingungen. Als Menge der Elementarereignisse wählen wir dann das Intervall [0, 1], als Menge der zufälligen Ereignisse die Klasse aller Borelschen Mengen, die aus Punkten dieses Intervalls gebildet werden können, und als Wahrscheinlichkeit das normierte Lebesguesche Maß, d. h., wir sehen als Wahrscheinlichkeit einer Borelschen Menge aus [0, 1] das Lebesguesche Maß dieser Menge an. Insbesondere ist also die Wahrscheinlichkeit des Intervalls $[0, e)$, wobei $0 < e \leqq 1$, gleich der Länge e dieses Intervalls. Die Zufallsvariable $X(e)$ definieren wir durch

$$X(e) = \inf_{F(y)=e} y \qquad (0 \leqq e \leqq 1).$$

Für ein gegebenes e ist also die Zufallsvariable $X(e)$ gleich der unteren Schranke aller derjenigen y, für welche $F(y) = e$ ist. Damit haben wir auch erreicht, daß die Verteilungsfunktionen der so definierten Zufallsvariablen $X(e)$ gleich der Funktion $F(x)$ ist, denn wir erhalten

$$P\big((X(e) < x)\big) = P\Big(\inf_{F(y)=e} y < x\Big) = F(x) \qquad (-\infty < x < \infty).$$

Wir haben also bewiesen:

Satz 2.2.1. *Eine reelle eindeutige Funktion $F(x)$ ist genau dann eine Verteilungsfunktion, wenn sie*

1. *nicht fallend und wenigstens linksseitig stetig ist,*
2. *die Bedingungen* (2.2.2) *erfüllt.*

Aus diesem Ergebnis folgt sofort: Die Verteilungsfunktion wird durch die Werte in ihren Stetigkeitsstellen bestimmt.

Wir wollen noch zeigen, daß die Menge H aller Unstetigkeitsstellen einer Verteilungsfunktion höchstens abzählbar ist. Dazu bezeichnen wir mit H_n die Menge aller derjenigen Punkte x, in denen die Verteilungsfunktion $F(x)$ einen Sprung größer oder gleich $\frac{1}{n}$ besitzt. Dann ist

$$H = H_1 \cup H_2 \cup \cdots.$$

Da für jedes n die Menge H_n endlich ist, ist H höchstens abzählbar.

2.3. Diskrete und stetige Zufallsvariable

A. Gewöhnlich beschäftigen wir uns mit zwei Arten von Zufallsvariablen: den diskreten und den stetigen Zufallsvariablen.

Definition 2.3.1. Eine Zufallsvariable, die mit der Wahrscheinlichkeit 1 Werte aus einer gewissen endlichen oder höchstens abzählbar unendlichen Menge S annimmt, wobei jeder Wert aus S eine positive Wahrscheinlichkeit hat, nennen wir *diskret*. Diese Werte nennen wir *Sprungstellen* und ihre Wahrscheinlichkeiten *Sprunghöhen*.

Beispiel 2.3.1. Eine Warenladung enthält fehlerlose und fehlerhafte Stücke. Wir betrachten zwei Elementarereignisse, die Wahl eines fehlerlosen und die eines fehlerhaften Exemplars. Die Wahrscheinlichkeit, ein fehlerloses Stück zufällig zu wählen, bezeichnen wir mit p, wobei wir voraussetzen, daß $0 < p < 1$ ist. Der Auswahl eines fehlerlosen Stückes aus dieser Ladung ordnen wir die Zahl Eins und der Auswahl eines fehlerhaften Stückes die Zahl Null zu. Wir erhalten eine diskrete Zufallsvariable, die mit positiver Wahrscheinlichkeit nur zwei Werte annehmen kann: 1 und 0, den ersten mit der Wahrscheinlichkeit p, den zweiten mit der Wahrscheinlichkeit $1 - p$.

Die diskrete Zufallsvariable X nehme die Werte x_i $(i = 1, 2, \ldots)$ mit den Wahrscheinlichkeiten p_i an. Für den Fall, daß die Anzahl der Sprungstellen endlich und gleich n ist, muß die Gleichung

$$\sum_{i=1}^{n} p_i = 1 \tag{2.3.1}$$

erfüllt sein, während, falls die Anzahl der Sprungstellen abzählbar ist, die Gleichung

$$\sum_{i=1}^{\infty} p_i = 1 \tag{2.3.1'}$$

gelten muß.

Die Definition 2.1.3 kann man in diesem Fall folgendermaßen formulieren:

Definition 2.3.2. Es sei x_i $(i = 1, 2, \ldots)$ eine beliebige Sprungstelle der diskreten Zufallsvariablen X. Die Wahrscheinlichkeit dafür, daß die Zufalls-

variable X den Wert x_i annimmt, nennen wir die *Wahrscheinlichkeitsfunktion* der diskreten Zufallsvariablen X und schreiben

$$P(X = x_i) = p_i, \tag{2.3.2}$$

wobei die Zahlen p_i $(i = 1, 2, \ldots)$ die Gleichung (2.3.1) bzw. (2.3.1′) erfüllen.

Die Wertemenge der Wahrscheinlichkeitsfunktion, die auf der Menge der Sprungstellen definiert ist, kann eine beliebige endliche oder abzählbare Menge von positiven Zahlen sein, die nur die Bedingung (2.3.1) bzw. (2.3.1′) zu erfüllen brauchen; sie kann auch aus einer einzigen Zahl bestehen. Dies ist der Fall, wenn die Zufallsvariable X nur einen einzigen Wert und den mit der Wahrscheinlichkeit 1 annimmt.

Die Verteilungsfunktion $F(x)$ hat hier die Form

$$F(x) = \sum_{x_i < x} p_i, \tag{2.3.3}$$

wobei die Summation über alle diejenigen Punkte x_i zu erstrecken ist, die die Ungleichung $x_i < x$ erfüllen.

Wir betrachten nun eine Zufallsvariable X, die Werte auf der x-Achse annimmt und keine Sprungstellen besitzt. Die Verteilungsfunktion einer solchen Zufallsvariablen ist stetig. Wir werden uns hauptsächlich mit der speziellen Klasse der *stetigen Zufallsvariablen* beschäftigen.

Definition 2.3.3. Wir nennen eine Zufallsvariable X *stetig*, wenn eine nichtnegative Funktion $f(x)$ existiert, die für jedes reelle x die Beziehung

$$F(x) = \int_{-\infty}^{x} f(t)\,dt \tag{2.3.4}$$

erfüllt, wobei $F(x)$ die Verteilungsfunktion der Zufallsvariablen X ist. Die Funktion $f(x)$ nennen wir die *Dichtefunktion* der Zufallsvariablen X.

Statt Dichtefunktion werden wir oft einfach *Dichte* sagen.

Die Dichtefunktion $f(x)$ genügt der Gleichung

$$F(\infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1. \tag{2.3.5}$$

Weiter gilt für beliebige a und b, wenn $a < b$ ist,

$$P(a \leqq X < b) = F(b) - F(a) = \int_{a}^{b} f(x)\,dx.$$

Weiterhin kann man beweisen, daß für eine beliebige Borelsche Menge S die Gleichung

$$P(S) = \int_{S} f(x)\,dx \tag{2.3.6}$$

besteht.

Ist die Dichtefunktion $f(x)$ im Punkt x stetig, so erhalten wir

$$F'(x) = f(x). \tag{2.3.7}$$

Für die Stetigkeitspunkte der Funktion $f(x)$ gilt also

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x \leqq X < x + \Delta x)}{\Delta x}. \tag{2.3.8}$$

Jede reelle, nichtnegative und auf der Zahlenachse $(-\infty, \infty)$ *integrierbare Funktion* $f(x)$, *die die Gleichung* (2.3.5) *erfüllt, ist Dichtefunktion einer gewissen stetigen Zufallsvariablen* X. Die durch die Formel

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

bestimmte Funktion $F(x)$ erfüllt nämlich, wovon man sich leicht überzeugt, alle Bedingungen für eine Verteilungsfunktion.

Beispiel 2.3.2. Auf der x-Achse definieren wir eine Dichtefunktion $f(x)$ folgendermaßen:

$$f(x) = \begin{cases} 0 & \text{für } x < 0, \\ \dfrac{x}{2} & \text{für } 0 \leqq x \leqq 2, \\ 0 & \text{für } x > 2. \end{cases}$$

Die Verteilungsfunktion der Zufallsvariablen X mit dieser Dichtefunktion hat die Gestalt

$$F(x) = \begin{cases} 0 & \text{für } x < 0, \\ \dfrac{x^2}{4} & \text{für } 0 \leqq x \leqq 2, \\ 1 & \text{für } x > 2. \end{cases}$$

B. Wir zeigen jetzt (siehe 1.3), daß daraus, daß die Wahrscheinlichkeit eines zufälligen Ereignisses gleich Null ist, noch nicht zu folgen braucht, daß das Ereignis unmöglich ist. Ebenso zeigen wir, daß daraus, daß die Wahrscheinlichkeit eines zufälligen Ereignisses gleich Eins ist, noch nicht das sichere Eintreffen des Ereignisses folgt.

Es sei X eine stetige Zufallsvariable mit der Wahrscheinlichkeitsdichte $f(x)$, und x_0 sei irgendein Punkt. Wie groß ist die Wahrscheinlichkeit dafür, daß $X = x_0$ ist?

Aus (2.3.6) erhalten wir

$$P(X = x_0) = \int_{x_0}^{x_0} f(x)\, dx = 0.$$

Trotzdem gehört der Punkt x_0 zur Menge der Elementarereignisse, die hier mit der x-Achse zusammenfällt.

Wir bezeichnen mit R' die x-Achse, aus der wir den Punkt x_0 entfernt haben. Es gilt die Gleichung

$$P(R') = \int_{R'} f(x)\,dx = 1,$$

obwohl das Ereignis, das darin besteht, daß X einen zu R' gehörigen Wert annimmt, kein sicher eintretendes Ereignis ist.

Wenn in der Verteilung der stetigen Zufallsvariablen X die Wahrscheinlichkeit irgendeines Ereignisses gleich Null ist, so kann man dieses Ereignis doch nicht als unmöglich (d. h. als unmöglich im Sinne der Definition 1.2.3) ansehen. Man betrachtet es als ein Ereignis, dessen Vorkommen sehr wenig wahrscheinlich ist. Ähnlich ist es, wenn in der Verteilung der stetigen Zufallsvariablen X die Wahrscheinlichkeit irgendeines Ereignisses gleich Eins ist. Ein solches Ereignis kann man als sehr wahrscheinlich ansehen, nicht aber als sicher (im Sinne der Definition 1.2.2).

2.4. Funktionen von Zufallsvariablen

A. Wir betrachten folgendes Beispiel:

Beispiel 2.4.1. Wir nehmen an, daß die Zufallsvariable X zwei Werte $x_1 = 5$ und $x_2 = 10$ mit den Wahrscheinlichkeiten $P(X = 5) = 1/3$ bzw. $P(X = 10) = 2/3$ annehmen kann. Wir transformieren X, indem wir

$$Y = 2X$$

setzen. Dann kann Y ebenfalls zwei Werte annehmen, nämlich $y_1 = 10$ und $y_2 = 20$. Dabei ist

$$P(Y = 10) = P(2X = 10) = P(X = 5) = \frac{1}{3},$$

$$P(Y = 20) = P(2X = 20) = P(X = 10) = \frac{2}{3}.$$

Die Zufallsvariable Y kann also die Werte $y_i = 2x_i$ $(i = 1, 2)$ mit derselben Wahrscheinlichkeit annehmen, mit der X die Werte x_i annimmt. Die Verteilungsfunktion der Zufallsvariablen X erhalten wir nach der Formel (2.3.3):

$$F(x) = \begin{cases} 0 & \text{für } x \leq 5, \\ \dfrac{1}{3} & \text{für } 5 < x \leq 10, \\ 1 & \text{für } x > 10. \end{cases}$$

Wir bezeichnen die Verteilungsfunktion der Zufallsvariablen Y mit $F_1(y)$ und erhalten

$$F_1(y) = P(Y < y) = P(2X < y) = P\left(X < \frac{y}{2}\right) = F\left(\frac{y}{2}\right),$$

also

$$F_1(y) = \begin{cases} 0 & \text{für } y \leqq 10, \\ \frac{1}{3} & \text{für } 10 < y \leqq 20, \\ 1 & \text{für } y > 20. \end{cases}$$

Wie wir sehen, kann man die Verteilung der Zufallsvariablen Y aus der Verteilung der Zufallsvariablen X erhalten.

Wenn allgemein $Y = g(X)$ eine eindeutige und stetige[1]) Transformation von X ist, dann ist Y ebenfalls eine Zufallsvariable, deren Verteilungsfunktion man aus der Verteilungsfunktion von X bestimmen kann.

Wir betrachten einige einfache Transformationen.

Es sei $F(x)$ die Verteilungsfunktion der Zufallsvariablen X. Wir betrachten die Transformation $Y = -X$.

Wir bezeichnen die Verteilungsfunktion von Y mit $F_1(y)$ und erhalten

$$F_1(y) = P(Y < y) = P(-X < y) = P(X > -y) = 1 - P(X \leqq -y). \tag{2.4.1}$$

Wenn X stetig ist, dann ist

$$F_1(y) = 1 - F(-y). \tag{2.4.2}$$

Wir schreiben die Gleichung (2.4.2) in der Gestalt

$$F_1(y) = 1 - \int_{-\infty}^{-y} f(x)\,dx,$$

bezeichnen die Dichte von Y mit $f_1(y)$ und erhalten

$$F_1'(y) = f_1(y) = f(-y). \tag{2.4.3}$$

Ist X diskret und der Punkt y eine Sprungstelle, dann erhalten wir

$$F_1(y) = 1 - F(-y) - P(X = y). \tag{2.4.4}$$

---

[1]) Es ist nicht nötig, die Funktion $g(X)$ als stetig vorauszusetzen. Es genügt anzunehmen, daß $g(X)$ eine Bairesche Funktion ist, d. h. eine eindeutige Funktion mit der Eigenschaft, daß die zu $g(X)$ inverse Funktion $h(Y)$ ein beliebiges Intervall der Form $(-\infty, y)$ in eine Borelsche Menge überführt, die wir hier mit $h(-\infty, y)$ bezeichnen. Dann ist nämlich die Verteilungsfunktion der Zufallsvariablen Y bestimmt; denn es ist

$$P(Y < y) = P(g(X) < y) = P(X \in h(-\infty, y)).$$

In Zukunft werden wir stets voraussetzen, daß die untersuchten Funktionen von Zufallsvariablen diese Eigenschaft haben.

Wir betrachten die allgemeine lineare Transformation

$$Y = aX + b$$

und unterscheiden zwei Fälle:

I. $a > 0$. Behalten wir die vorherigen Bezeichnungen bei, so ist

$$F_1(y) = P(aX + b < y) = P\left(X < \frac{y-b}{a}\right) = F\left(\frac{y-b}{a}\right). \tag{2.4.5}$$

Die Gleichung (2.4.5) gilt für diskrete und stetige Zufallsvariable. Im letzten Fall gilt außerdem

$$F_1'(y) = f_1(y) = \frac{1}{a} f\left(\frac{y-b}{a}\right). \tag{2.4.6}$$

II. $a < 0$. In diesem Fall erhalten wir

$$F_1(y) = P(aX + b < y) = P\left(X > \frac{y-b}{a}\right) = 1 - P\left(X \leqq \frac{y-b}{a}\right). \tag{2.4.7}$$

Wenn X stetig ist, folgt aus der Formel (2.4.7)

$$F_1(y) = 1 - F\left(\frac{y-b}{a}\right), \tag{2.4.8}$$

$$F_1'(y) = f_1(y) = -\frac{1}{a} f\left(\frac{y-b}{a}\right). \tag{2.4.9}$$

Die Formeln (2.4.6) und (2.4.9) kann man zu einer Formel zusammenfassen, nämlich

$$F_1'(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right). \tag{2.4.10}$$

Die Formel (2.4.3) ist ein Spezialfall der Formel (2.4.10).

Ist X diskret, so erhalten wir aus der Formel (2.4.7)

$$F_1(y) = 1 - F\left(\frac{y-b}{a}\right) - P\left(X = \frac{y-b}{a}\right), \tag{2.4.11}$$

wenn der Punkt $\frac{y-b}{a}$ eine Sprungstelle der Zufallsvariablen X ist. In den übrigen Punkten ist $P\left(X = \frac{y-b}{a}\right) = 0$, und die Formel (2.4.11) stimmt mit der Formel (2.4.8) überein.

B. Weiter sei X eine Zufallsvariable mit der Verteilungsfunktion $F(x)$. Wir betrachten die Transformation $Y = X^2$. Die Zufallsvariable Y nimmt keine negativen Werte an. $F_1(x)$ sei die Verteilungsfunktion der Zufallsvariablen Y; dann erhalten wir

$$F_1(y) = \begin{cases} 0 & \text{für } y \leqq 0, \\ P(Y < y) = P(X^2 < y) = P\left(-\sqrt{y} < X < \sqrt{y}\right) & \text{für } y > 0. \end{cases} \tag{2.4.12}$$

Ist die Zufallsvariable X stetig, so ergibt sich aus (2.4.12)

$$F_1(y) = \begin{cases} 0 & \text{für } y \leqq 0, \\ F\left(\sqrt{y}\right) - F\left(-\sqrt{y}\right) & \text{für } y > 0. \end{cases} \tag{2.4.13}$$

Wenn $f(x)$ die Dichte der Zufallsvariablen X ist, dann erhalten wir die Dichte $f_1(y)$ der Zufallsvariablen Y aus der Formel (2.4.13):

$$f_1(y) = \begin{cases} 0 & \text{für } y \leqq 0, \\ \dfrac{F'\left(\sqrt{y}\right) + F'\left(-\sqrt{y}\right)}{2\sqrt{y}} = \dfrac{f\left(\sqrt{y}\right) + f\left(-\sqrt{y}\right)}{2\sqrt{y}} & \text{für } y > 0. \end{cases} \tag{2.4.14}$$

Ist aber X diskret, dann erhalten wir aus (2.4.12)

$$F_1(y) = \begin{cases} 0 & \text{für } y \leqq 0, \\ F\left(\sqrt{y}\right) - F\left(-\sqrt{y}\right) - P\left(X = -\sqrt{y}\right) & \text{für } y > 0. \end{cases} \tag{2.4.15}$$

Wenn der Punkt $-\sqrt{y}$ keine Sprungstelle der Zufallsvariablen X ist, dann ist $P\left(X = -\sqrt{y}\right) = 0$, und (2.4.15) stimmt mit (2.4.13) überein.

Es seien $x_1, x_2, \ldots$ die Sprungstellen der Zufallsvariablen X und $y_1, y_2, \ldots$ die ihnen auf Grund der Beziehung $y_i = x_i^2$ entsprechenden Punkte. Dann ist

$$P(Y = y_i) = P(X^2 = y_i) = P\left(X = -\sqrt{y_i}\right) + P\left(X = \sqrt{y_i}\right). \tag{2.4.16}$$

Beispiel 2.4.2. Wir nehmen an, daß die Zufallsvariable X nur zwei Werte annehmen kann, nämlich $x_1 = -1$ und $x_2 = 1$; dabei sei $P(X = -1) = P(X = 1) = 1/2$. Wir setzen $Y = X^2$. Die Zufallsvariable Y kann dann nur einen Wert annehmen, und es ist

$$P(Y = 1) = P(X = -1) + P(X = 1) = \frac{1}{2} + \frac{1}{2} = 1.$$

C. Es sei X eine stetige Zufallsvariable mit der Dichte $f(x)$. Wir betrachten die eineindeutige Transformation, die durch eine stetig differenzierbare Funktion $y = g(x)$ gegeben wird.

Es sei $g'(x) \neq 0$ in dem Intervall $[x_1, x_2)$ und $y_1 = g(x_1)$, $y_2 = g(x_2)$. Mit $x = h(y)$ bezeichnen wir die Umkehrfunktion von $g(x)$. Auf Grund der Voraussetzung ist die Funktion $h(y)$ eindeutig, und ihre Ableitung $h'(y)$ ist im Intervall $[y_1, y_2)$ endlich und stetig. Wir erhalten

$$P(x_1 \leqq X < x_2) = \int_{x_1}^{x_2} f(x)\,dx = \int_{y_1}^{y_2} f[h(y)]\,h'(y)\,dy. \tag{2.4.17}$$

Ist die Ableitung $h'(y)$ positiv, so folgt $y_1 < y_2$. Das Integral auf der rechten Seite von (2.4.17) gibt dann die Wahrscheinlichkeit $P(y_1 \leqq Y < y_2)$ an, wobei $Y = g(X)$; ist aber $h'(y) < 0$, so folgt $y_2 < y_1$, und Formel (2.4.17) ergibt

$$\begin{aligned} P(x_1 \leqq X < x_2) &= -\int_{y_2}^{y_1} f[h(y)]\,h'(y)\,dy \\ &= \int_{y_2}^{y_1} f[h(y)]\,|h'(y)|\,dy = P(y_2 \leqq Y < y_1). \end{aligned} \tag{2.4.18}$$

Aus (2.4.17) und (2.4.18) folgt, daß die Zufallsvariable $Y = g(X)$ die Wahrscheinlichkeitsdichte

$$f[h(y)]\,|h'(y)| \tag{2.4.19}$$

hat. Die Formel (2.4.10) ist ein Spezialfall von (2.4.19).

2.5. Mehrdimensionale Zufallsvariable

A. Das nachstehende Beispiel soll den Begriff einer mehrdimensionalen Zufallsvariablen erläutern.

Beispiel 2.5.1. Die Tabelle 2.5.1 enthält Angaben über die Zusammensetzung der Bevölkerung Polens in Hinsicht auf Geschlecht und Alter. Die Angaben stammen aus der allgemeinen Volkszählung vom Jahre 1931. Jeder Einwohner Polens ist in der Tafel nach zwei Kennzeichen eingeordnet, nach seinem Geschlecht und seinem Lebensalter. Diesen zwei Kennzeichen kann man Zahlenwerte zuordnen. Bei der Bearbeitung der Ergebnisse der Volkszählung fertigt man auf besonderen Maschinen (Perforiermaschinen) für jede von der Zählung erfaßte Person eine Lochkarte an. Auf diesen Karten ordnet man den untersuchten Kennzeichen gewisse Zahlen zu. So gibt man z. B. Männern gewöhnlich die Zahl Eins und Frauen die Zahl Null. Ähnlich erhält jede Altersgruppe eine bestimmte Zahl.

In jeder Karte sind also zwei Zahlen gelocht, die einer bestimmten Altersgruppe und einem bestimmten Geschlecht entsprechen. Betrachten wir das Herausgreifen einer solchen Karte, d. h. die Zugehörigkeit einer Person zu einer bestimmten Alters- und Geschlechtsgruppe, als Elementarereignis, so ist jedem Elementarereignis ein Zahlenpaar zugeordnet. Hierbei handelt es sich um eine sogenannte zweidimensionale Zufallsvariable.

Wir betrachten die Menge E der Elementarereignisse. Auf dem Borelschen Mengenkörper Z, welcher aus Teilmengen A der Menge E der Elementarereignisse gebildet ist, sei die Wahrscheinlichkeit $P(A)$ gegeben. Wir definieren auf der Menge E ein System von n reellen eindeutigen Funktionen, die jedem

Tabelle 2.5.1.*) *Die Bevölkerung Polens, angeordnet nach Geschlecht und Alter. Als Grundlage diente die Volkszählung aus dem Jahre 1931*

Altersgruppe	Männer (in 1000)	Frauen (in 1000)
0—4	2020	1962
5—9	2005	1962
10—14	1405	1372
15—19	1474	1562
20—29	2931	3213
30—39	1999	2255
40—49	1391	1596
50—59	1052	1201
60—69	753	875
70 und älter	386	474
Gesamtzahl	15416	16472

*) Quelle: Mały Rocznik Statystyczny 1939, dział II, tablica 11.

Elementarereignis ein System von n Zahlen $x_1, x_2, \ldots, x_n$ zuordnen. Mit anderen Worten, wir ordnen jedem Elementarereignis der Menge E einen Punkt im n-dimensionalen euklidischen Raum zu.

Definition 2.5.1. Wir nennen ein System von n reellen und eindeutigen Funktionen $(X_1, X_2, \ldots, X_n)$, die auf der Menge E definiert sind, eine *n-dimensionale Zufallsvariable*, wenn das Urbild A eines jeden n-dimensionalen Intervalls[1]) I der Form $(-\infty, \ldots, -\infty; a_1, \ldots, a_n)$ ein zufälliges Ereignis ist.

Für jedes Intervall I definieren wir die Wahrscheinlichkeit[2]) $P^{(x)}(I)$ dafür, daß die n-dimensionale Zufallsvariable einen Wert annimmt, der zum Intervall I gehört, indem wir $P^{(x)}(I) = P(A)$ setzen, wenn A das Urbild von I ist.

Ähnlich wie in 2.1 kann man zeigen, daß dadurch die Wahrscheinlichkeit $P^{(x)}(S)$ dafür erklärt ist, daß die Zufallsvariable $(X_1, X_2, \ldots, X_n)$ Werte $(x_1, x_2, \ldots, x_n)$ aus einer beliebigen Borelschen Menge S des n-dimensionalen euklidischen Raumes annimmt, d. h., daß die Wahrscheinlichkeitsverteilung der Zufallsvariablen $(X_1, X_2, \ldots, X_n)$ dadurch bestimmt ist.

---

[1]) Unter einem n-dimensionalen Intervall I im n-dimensionalen euklidischen Raum verstehen wir die Menge derjenigen Punkte, die die Eigenschaft haben, daß ihre i-ten Koordinaten $(i = 1, 2, \ldots, n)$ in einem gewöhnlichen Intervall $a_i < x_i < b_i$ enthalten sind. Unter dem Intervall $(-\infty, \ldots, -\infty; a_1, \ldots, a_n)$ verstehen wir also die Punktmenge im n-dimensionalen euklidischen Raum, für die die Ungleichungen $-\infty < x_i < a_i$ $(i = 1, 2, \ldots, n)$ gelten.

[2]) Vgl. den Schluß von 2.1.

Die Funktion $P^{(x)}(S)$ wird wieder *Wahrscheinlichkeitsfunktion* genannt. Statt *mehrdimensionale Zufallsvariable* sagt man auch *Zufallsvektor*.

B. Wir werden uns in den weiteren Darlegungen meistens auf zweidimensionale Zufallsvariable beschränken.

Definition 2.5.2. Die reelle Funktion $F(x, y)$, die durch

$$F(x, y) = P(X < x, Y < y) \tag{2.5.1}$$

definiert wird, nennen wir die *Verteilungsfunktion der zweidimensionalen Zufallsvariablen* (X, Y).

Wie in 2.2 kann man zeigen, daß die Verteilungsfunktion $F(x, y)$ in jedem der Argumente x und y eine nichtfallende und wenigstens linksseitig stetige Funktion ist und daß sie die folgenden Bedingungen erfüllt:

$$F(-\infty, y) = F(x, -\infty) = 0, \quad F(\infty, \infty) = 1. \tag{2.5.2}$$

Wir bemerken, daß auch die folgende Gleichung erfüllt ist:

$$\begin{aligned} P(x_1 \leqq X < x_2, y_1 \leqq Y < y_2) &= P(X < x_2, Y < y_2) - P(X < x_2, Y < y_1) \\ &\quad - P(X < x_1, Y < y_2) + P(X < x_1, Y < y_1) \\ &= F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1). \end{aligned} \tag{2.5.3}$$

In Abb. 2.5.1 ist die Formel (2.5.3) erläutert.

Dafür, daß eine reelle Funktion $F(x, y)$ die Verteilungsfunktion einer zweidimensionalen Zufallsvariablen ist, ist es nicht hinreichend, daß die Funktion

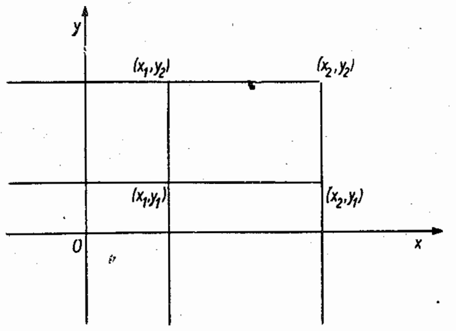


Abb. 2.5.1

1. nichtfallend und linksseitig stetig in jeder Zufallsvariablen ist und daß sie 2. die Bedingungen (2.5.2) erfüllt.

Es muß außerdem, wie man aus (2.5.3) erkennt, für alle Paare x_2, x_1 mit $x_2 > x_1$ und y_2, y_1 mit $y_2 > y_1$ die Bedingung

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geqq 0$$

erfüllt sein.

Wie das folgende Beispiel zeigt, sind die beiden zuerst angegebenen Bedingungen nicht hinreichend:

Es sei

$$F(x, y) = \begin{cases} 0 & \text{für} \quad x + y \leqq 0, \\ 1 & \text{für} \quad x + y > 0. \end{cases}$$

Die Funktion $F(x, y)$ nimmt also den Wert 0 in den Punkten auf und unterhalb der Geraden $y = -x$ an, während sie den Wert 1 in den Punkten oberhalb dieser Geraden annimmt. Die so definierte Funktion $F(x, y)$ ist nichtfallend und linksseitig stetig in jeder Zufallsvariablen x und y und erfüllt die Bedingungen (2.5.2). Dennoch ist $F(x, y)$ keine Verteilungsfunktion einer Wahrscheinlichkeitsverteilung.

Betrachten wir nämlich irgendein Rechteck in der x, y-Ebene, von dem drei Ecken oberhalb und eine Ecke unterhalb der Geraden $y = -x$ liegen, zum Beispiel das Rechteck mit den Ecken (3, 3), (3, −1), (−1, 3), (−1, −1). Wenn wir die Formel (2.5.3) anwenden, erhalten wir

$$P(-1 \leqq X < 3, -1 \leqq Y < 3) = F(3, 3) - F(3, -1) - F(-1, 3) + F(-1, -1) = -1.$$

Das ist jedoch unmöglich, da eine Wahrscheinlichkeit nicht negativ sein kann.

Es gilt der folgende Satz:

Satz 2.5.1. *Eine reelle eindeutige Funktion $F(x, y)$ ist genau dann Verteilungsfunktion einer zweidimensionalen Zufallsvariablen, wenn sie*

1. *in jedem der Argumente x und y nichtfallend und wenigstens linksseitig stetig ist,*
2. *die Bedingungen* (2.5.2) *erfüllt,*
3. *für beliebige Punkte (x_1, y_1), (x_2, y_2), wobei $x_1 < x_2$, $y_1 < y_2$ ist, der Ungleichung*

$$F(x_2, x_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geqq 0$$

genügt.

C. Wir wollen nun mehrdimensionale diskrete und stetige Zufallsvariable betrachten.

Definition 2.5.3. Die zweidimensionale Zufallsvariable (X, Y) ist *diskret*, wenn sie mit der Wahrscheinlichkeit 1 Wertepaare annimmt, die zu einer gewissen höchstens abzählbaren Menge S von Paaren gehören, wobei jedes Paar (x_i, y_k)

eine positive Wahrscheinlichkeit p_{ik} besitzt. Diese Wertepaare nennen wir *Sprungstellen* und ihre Wahrscheinlichkeit *Sprunghöhe*.

Die Indizes i und k durchlaufen gewisse Mengen von ganzen Zahlen; dabei braucht aber die Wertemenge von k nicht für alle Indizes i die gleiche zu sein, wie auch die Wertemenge von i nicht für alle Indizes k die gleiche zu sein braucht.

Aus dem zugrunde gelegten Axiomensystem folgt die Gleichung

$$\sum_i \sum_k p_{ik} = 1. \tag{2.5.4}$$

Die Verteilungsfunktion $F(x, y)$ hat hier die Form

$$F(x, y) = \sum_{\substack{x_i < x \\ y_k < y}} p_{ik}; \tag{2.5.5}$$

dabei ist die Summation über alle Punkte (x_i, y_k) zu erstrecken, die die Ungleichungen $x_i < x$ und $y_k < y$ erfüllen.

Die Definition der Wahrscheinlichkeitsfunktion kann man in diesem Fall auf die folgende Form bringen:

Definition 2.5.4. Es sei (x_i, y_k) $(i = 1, 2, \ldots; k = 1, 2, \ldots)$ eine beliebige Sprungstelle der diskreten Zufallsvariablen (X, Y). Die Wahrscheinlichkeit, daß die Zufallsvariable (X, Y) das Wertepaar (x_i, y_k) annimmt, nennen wir *Wahrscheinlichkeitsfunktion der Zufallsvariablen* (X, Y) und schreiben

$$P(X = x_i, Y = y_k) = p_{ik}. \tag{2.5.6}$$

Jetzt wollen wir den Begriff einer stetigen zweidimensionalen Zufallsvariablen definieren.

Definition 2.5.5. Die zweidimensionale Zufallsvariable (X, Y) heißt *stetig*, wenn es eine nichtnegative Funktion $f(x, y)$ gibt, die für jedes reelle Zahlenpaar (x, y) die Beziehung

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)\, dy\, dx \tag{2.5.7}$$

erfüllt; dabei ist $F(x, y)$ die Verteilungsfunktion der Zufallsvariablen (X, Y). Die Funktion $f(x, y)$ nennt man *Dichtefunktion*.

Die Dichtefunktion $f(x, y)$ genügt der Beziehung

$$F(\infty, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1. \tag{2.5.8}$$

Ist die Dichtefunktion $f(x, y)$ im Punkt (x, y) stetig, dann gilt die Beziehung

$$\frac{\partial^2 F(x, y)}{\partial x\, \partial y} = f(x, y). \tag{2.5.9}$$

Also erhalten wir für einen Stetigkeitspunkt von $f(x, y)$

$$f(x, y) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{P(x \leqq X < x + \Delta x, y \leqq Y < y + \Delta y)}{\Delta x\, \Delta y}. \tag{2.5.10}$$

Alle in diesem Paragraphen eingeführten Begriffe für zweidimensionale Zufallsvariable kann man leicht auf n-dimensionale ($n \geqq 3$) Zufallsvariable verallgemeinern. Insbesondere ist die Verteilungsfunktion einer n-dimensionalen Zufallsvariablen $(X_1, X_2, \ldots, X_n)$ durch die folgende Formel definiert:

$$F(x_1, x_2, \ldots, x_n) = P(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n). \tag{2.5.11}$$

Die Verallgemeinerung von (2.5.3) auf den Fall der n-dimensionalen Zufallsvektoren $(X_1, X_2, \ldots, X_n)$ ist die Formel

$$\begin{aligned} &P(x_1 \leqq X_1 < x_1 + h_1, \ldots, x_n \leqq X_n < x_n + h_n) \\ &= F(x_1 + h_1, \ldots, x_n + h_n) \\ &\quad - \sum_{i=1}^{n} F(x_1 + h_1, \ldots, x_{i-1} + h_{i-1}, x_i, x_{i+1} + h_{i+1}, \ldots, x_n + h_n) \\ &\quad + \sum_{\substack{i,j=1 \\ i<j}}^{n} F(x_1 + h_1, \ldots, x_i, \ldots, x_j, \ldots, x_n + h_n) + \cdots \\ &\quad + (-1)^n F(x_1, \ldots, x_n), \end{aligned} \tag{2.5.12}$$

wobei $h_i > 0$ $(i = 1, 2, \ldots, n)$ gilt.

Der Satz 2.5.1 läßt sich ebenfalls auf n-dimensionale Zufallsvariable verallgemeinern. Anstelle der Ungleichung (2.5.4) hätten wir dann die Ungleichung, der zufolge die rechte Seite von (2.5.12) nicht negativ ist.

Da die Verteilungsfunktion $F(x_1, \ldots, x_n)$ linksseitig stetig und bezüglich einer jeden unabhängigen Variablen nicht fallend ist, folgt hieraus, daß die Werte von $F(x_1, \ldots, x_n)$ an ihren Stetigkeitsstellen die Funktion vollständig bestimmen.

Wir wollen noch einen wesentlichen Unterschied zwischen der eindimensionalen und den mehrdimensionalen Verteilungsfunktionen hervorheben, und zwar: Besitzt eine eindimensionale Zufallsvariable X keine Sprungstellen, so ist ihre Verteilungsfunktion $F(x)$ überall stetig. Demgegenüber kann die Verteilungsfunktion $F(x_1, \ldots, x_n)$ Unstetigkeitsstellen besitzen, obwohl die Zufallsvariable $(X_1, \ldots, X_n)$ keine Sprungstellen aufweist. Dies ist dann der Fall, wenn eine

durch die Gleichungen $X_{j_1} = a_1, \ldots, X_{j_r} = a_r$ definierte Hyperebene (hierbei gilt $1 \leqq r < n$; $a_1, \ldots, a_r$ sind Konstanten) derart existiert, daß

$$P(X_{j_1} = a_1, \ldots, X_{j_r} = a_r) > 0$$

gilt.

Beispiel 2.5.2. Wir betrachten einen Zufallsvektor (X, Y) mit einer Verteilungsfunktion $F(x,y)$ der Form

$$F(x,y) = \begin{cases} 0 & \text{in den Gebieten } (-\infty < x \leqq 0,\ -\infty < y < \infty) \\ & \text{und } (0 < x < \infty,\quad -\infty < y \leqq 0), \\ xy & \text{im Gebiet } \left(0 \leqq x \leqq 1,\quad 0 \leqq y \leqq \frac{1}{2}\right), \\ \frac{x}{2} & \text{im Gebiet } \left(0 \leqq x \leqq 1,\ \frac{1}{2} \leqq y < \infty\right), \\ y & \text{im Gebiet } (1 < x < \infty,\quad 0 \leqq y \leqq 1), \\ 1 & \text{im Gebiet } (x > 1,\quad y > 1). \end{cases}$$

Der Zufallsvektor (X, Y) fällt mit der Wahrscheinlichkeit $\frac{1}{2}$ in einen Punkt von $(x = 1, \frac{1}{2} \leqq y \leqq 1)$. Offenbar ist jeder Punkt mit den Koordinaten $(1,y)$ mit $\frac{1}{2} < y < \infty$ ein Unstetigkeitspunkt der Verteilungsfunktion $F(x,y)$, obwohl der Zufallsvektor (Y, X) keine Sprungstellen besitzt.

Wenn alle „Eckpunkte" des durch die Ungleichungen

$$x_1 \leqq X_1 < x_1 + h_1, \quad x_2 \leqq X_2 < x_2 + h_2, \quad \ldots, \quad x_n \leqq X_n < x_n + h_n$$

gegebenen verallgemeinerten Intervalls Stetigkeitspunkte der Verteilungsfunktion $F(x_1, \ldots, x_n)$ sind, gilt für die „Oberfläche" S des Intervalls

$$P[(X_1, X_2, \ldots, X_n) \in S] = 0. \tag{2.5.13}$$

Definition 2.5.6. Ein Intervall im verallgemeinerten oder üblichen Sinne heißt *Stetigkeitsintervall*, wenn für dieses die Bedingung (2.5.13) erfüllt ist.

Es sei darauf hingewiesen, daß ein Intervall durchaus ein Stetigkeitsintervall sein kann, obwohl die Verteilungsfunktion in einigen „Eckpunkten" dieses Intervalls unstetig sein kann. So ist im Beispiel 2.5.2 das Rechteck mit den Eckpunkten (1, 2), (1, 3), (2, 2) und (2, 3) ein Stetigkeitsintervall im Sinne der Definition 2.5.6, obwohl die Eckpunkte (1, 2) und (1, 3) Unstetigkeitsstellen der Verteilungsfunktion $F(x, y)$ sind.

Ferner sei bemerkt, daß das verallgemeinerte Intervall der Form $(-\infty, -\infty, \ldots, -\infty, a_1, a_2, \ldots, a_n)$ dann und nur dann ein Stetigkeitsintervall ist, wenn der Punkt $(a_1, a_2, \ldots, a_n)$ ein Stetigkeitspunkt der Verteilungsfunktion ist.

2.6. Randverteilungen

Es sei (X, Y) eine diskrete Zufallsvariable, die die Werte (x_i, y_k) annimmt. Wir schreiben folgendermaßen:

$$P(X = x_i, Y = y_k) = p_{ik},$$

$$p_{\cdot k} = \sum_i p_{ik}, \tag{2.6.1}$$

$$p_{i\cdot} = \sum_k p_{ik}. \tag{2.6.2}$$

Dann ist

$$p_{\cdot k} = \sum_i p_{ik} = P(X = x_1, Y = y_k) + P(X = x_2, Y = y_k) + P(X = x_3, Y = y_k) + \cdots;$$

also ist $p_{\cdot k}$ die Wahrscheinlichkeit dafür, daß $Y = y_k$ ist, wenn X irgendeinen der möglichen Werte annimmt. Deshalb kann man

$$p_{\cdot k} = P(Y = y_k)$$

schreiben. Offensichtlich ist ferner

$$\sum_k p_{\cdot k} = \sum_i \sum_k p_{ik} = 1.$$

Die Zahlen $p_{\cdot k}$ stellen also die Wertemengen einer Wahrscheinlichkeitsfunktion dar. Die durch diese Wahrscheinlichkeitsfunktion bestimmte Verteilung nennen wir *Randverteilung* der Zufallsvariablen Y in der zweidimensionalen Verteilung der Variablen (X, Y). Ebenso bilden die Zahlen $p_{i\cdot}$ die Wertemenge der Randwahrscheinlichkeitsfunktion, der Zufallsvariablen X in der zweidimensionalen Verteilung der Variablen (X, Y).

Beispiel 2.6.1. In einer Urne befinden sich 21 Papierstreifen. Jeder Streifen ist mit einer der natürlichen Zahlen 1, 2, ..., 21 versehen, wobei jede dieser Zahlen nur einmal auftritt. Uns interessieren zwei Merkmale dieser Zahlen: die Teilbarkeit durch 2 und durch 3. Wir ziehen einen Streifen. Dem Auftreten einer geraden Zahl ordnen wir die Zahl Eins zu, dem Auftreten einer ungeraden Zahl die Zahl Null und bezeichnen diese Zufallsvariable mit X. Die Variable X nimmt also die beiden Werte $x_1 = 1$ und $x_2 = 0$ an. Ebenso nehme die Variable Y den Wert $y_1 = 1$ beim Auftreten einer durch 3 teilbaren Zahl und den Wert $y_2 = 0$ beim Auftreten einer nicht durch 3 teilbaren Zahl an.

Unter den 21 Zahlen gibt es

3 Zahlen, die durch 2 und durch 3 teilbar sind,
7 Zahlen, die durch 2, aber nicht durch 3 teilbar sind,
4 Zahlen, die nicht durch 2, aber durch 3 teilbar sind,
7 Zahlen, die nicht durch 2 und nicht durch 3 teilbar sind.

Setzen wir voraus, daß die Wahrscheinlichkeit, gezogen zu werden, für alle Streifen dieselbe ist, so erhalten wir

$$p_{11} = P(X = 1,\ Y = 1) = \frac{3}{21}, \qquad p_{21} = P(X = 0,\ Y = 1) = \frac{4}{21},$$

$$p_{12} = P(X = 1,\ Y = 0) = \frac{7}{21}, \qquad p_{22} = P(X = 0,\ Y = 0) = \frac{7}{21}.$$

Wir können zwei Randverteilungen betrachten: eine, wenn wir die Zahlen ausschließlich nach ihrer Teilbarkeit durch 2, die andere, wenn wir sie ausschließlich nach ihrer Teilbarkeit durch 3 klassifizieren.

Die Gleichung (2.6.1) liefert

$$p_{\cdot 1} = p_{11} + p_{21} = \frac{3}{21} + \frac{4}{21} = \frac{7}{21},$$

$$p_{\cdot 2} = p_{12} + p_{22} = \frac{7}{21} + \frac{7}{21} = \frac{14}{21}.$$

Das ist die Randverteilung bezüglich der Teilbarkeit durch 3. Ebenso erhalten wir aus (2.6.2)

$$p_{1\cdot} = p_{11} + p_{12} = \frac{3}{21} + \frac{7}{21} = \frac{10}{21},$$

$$p_{2\cdot} = p_{21} + p_{22} = \frac{4}{21} + \frac{7}{21} = \frac{11}{21}.$$

Das ist die Randverteilung bezüglich der Teilbarkeit durch 2.

Die Bezeichnung Randverteilung wird verständlich, wenn wir dieses Beispiel in Form der Tabelle 2.6.1 anordnen.

Tabelle 2.6.1. *Die Anzahl der Zahlen in den einzelnen Gruppen*

	durch 2 teilbare Zahlen	durch 2 nicht teilbare Zahlen	insgesamt
durch 3 teilbare Zahlen	3	4	7
durch 3 nicht teilbare Zahlen	7	7	14
insgesamt	10	11	21

Wie wir sehen, erhalten wir in der untersten Zeile der Tabelle 2.6.1 die Verteilung hinsichtlich der Teilbarkeit durch 2 und in der rechten Außenspalte die Verteilung hinsichtlich der Teilbarkeit durch 3.

Es sei allgemein $F(x, y)$ die Verteilungsfunktion der zweidimensionalen Zufallsvariablen (X, Y). Die Verteilungsfunktion der Randverteilung von X hat die Form

$$F(x, \infty) = P(X < x,\ Y < \infty). \tag{2.6.3}$$

Wenn (X, Y) eine diskrete Zufallsvariable ist, dann nimmt (2.6.3) die Gestalt

$$F(x, \infty) = \sum_{x_i < x} \sum_{y_k} p_{ik} \tag{2.6.4}$$

an, wobei die Summation über alle k und über diejenigen i zu erstrecken ist, für die die Ungleichung $x_i < x$ erfüllt ist.

Ist (X, Y) eine stetige Zufallsvariable, so erhalten wir

$$F(x, \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx. \tag{2.6.5}$$

In der zweidimensionalen Verteilung von (X, Y) hat die Dichte der Randverteilung von X die Gestalt

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy. \tag{2.6.6}$$

Ähnliche Formeln kann man für die Randverteilung der Zufallsvariablen Y erhalten.

Es sei $F(x_1, x_2, \ldots, x_n)$ die Verteilungsfunktion des Vektors $(X_1, X_2, \ldots, X_n)$ für $n > 2$. Man kann dann $\binom{n}{k}$ k-dimensionale Randverteilungen erhalten $(k = 1, 2, \ldots, n - 1)$. So hat zum Beispiel die Randverteilungsfunktion der Zufallsvariablen (X_1, X_2) die Form

$$F(x_1, x_2, \infty, \ldots, \infty) = P(X_1 < x_1, X_2 < x_2, X_3 < \infty, \ldots, X_n < \infty).$$

2.7. Bedingte Verteilungen

A. In 1.5 haben wir die bedingte Wahrscheinlichkeit $P(A \mid B)$ definiert; das war die Wahrscheinlichkeit des Ereignisses A unter der Bedingung B. Wir wollen uns nun mit den bedingten Verteilungen beschäftigen.

Es sei (X, Y) eine zweidimensionale Zufallsvariable, wobei X die Werte x_i $(i = 1, 2, \ldots)$ und Y die Werte y_k $(k = 1, 2, \ldots)$ annehmen kann. Es sei weiter

$$P(X = x_i, Y = y_k) = p_{ik}.$$

Die Randverteilungen sind dann

$$P(X = x_i) = p_{i\cdot} = \sum_k p_{ik},$$

$$P(Y = y_k) = p_{\cdot k} = \sum_i p_{ik}.$$

Für jedes i und jedes k definieren wir die Wahrscheinlichkeiten

$$P(X = x_i \mid Y = y_k) = \frac{p_{ik}}{p_{\cdot k}} \tag{2.7.1}$$

und

$$P(Y = y_k | X = x_i) = \frac{p_{ik}}{p_{i\cdot}}. \tag{2.7.2}$$

Bei festem y_k und veränderlichem, alle möglichen Sprungstellen durchlaufendem x_i stellt der Ausdruck (2.7.1) die *bedingte Wahrscheinlichkeitsfunktion* der diskreten Zufallsvariablen X unter der Bedingung $Y = y_k$ dar. Bei festem x_i und veränderlichem, alle möglichen Sprungstellen durchlaufendem y_k stellt der Ausdruck (2.7.2) die *bedingte Wahrscheinlichkeitsfunktion* der diskreten Zufallsvariablen Y unter der Bedingung $X = x_i$ dar.

In der Tat sind die Ausdrücke auf den rechten Seiten der Gleichungen (2.7.1) und (2.7.2) nichtnegative Zahlen und höchstens gleich Eins. Außerdem bestehen, wie man leicht nachprüft, die Gleichungen

$$\sum_i P(X = x_i | Y = y_k) = \frac{\sum_i p_{ik}}{p_{\cdot k}} = \frac{p_{\cdot k}}{p_{\cdot k}} = 1,$$

$$\sum_k P(Y = y_k | X = x_i) = \frac{\sum_k p_{ik}}{p_{i\cdot}} = \frac{p_{i\cdot}}{p_{i\cdot}} = 1.$$

B. Es sei $f(x, y)$ die Dichte einer zweidimensionalen stetigen Zufallsvariablen (X, Y). Wir betrachten das Intervall $[x, x + h)$ und das Ereignis $x \leqq X < x + h$; dabei sei

$$P(x \leqq X < x + h) > 0.$$

Für jeden Wert y und für jedes Intervall $[x, x + h]$ definieren wir die bedingte Wahrscheinlichkeit

$$P(Y < y | x \leqq X < x + h) = \frac{P(Y < y, x \leqq X < x + h)}{P(x \leqq X < x + h)}$$
$$= \frac{\int\limits_x^{x+h} \int\limits_{-\infty}^{y} f(x,y)\, dy\, dx}{\int\limits_x^{x+h} \int\limits_{-\infty}^{\infty} f(x,y)\, dy\, dx}. \tag{2.7.3}$$

Der Ausdruck (2.7.3) ist die bedingte Verteilungsfunktion der Zufallsvariablen Y unter der Bedingung, daß die Zufallsvariable X die Ungleichung $x \leqq X < x + h$ erfüllt; dabei wird $P(x \leqq X < x + h) > 0$ vorausgesetzt. Der Ausdruck auf der rechten Seite der Formel (2.7.3) ist nämlich bei festem x und h eine stetige und nichtfallende Funktion von y, die die Werte 0 für $y = -\infty$ und 1 für $y = \infty$ annimmt. Schwierigkeiten treten auf, wenn wir die bedingte Wahrscheinlichkeit $P(Y < y | X = x)$ angeben wollen, da in diesem Fall $P(X = x) = 0$ ist. Wir

beschränken uns hier auf den Fall, daß der Grenzwert des Ausdrucks (2.7.3) für $h \to 0$ existiert. (Vgl. 3.6.C und den Anhang.)

Wir nehmen an, daß die Dichtefunktion $f(x,y)$ überall stetig und die Randdichtefunktion

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \tag{2.7.4}$$

eine stetige Funktion von x ist; außerdem sei $f_1(x)$ im betrachteten Punkt x positiv. Teilt man Zähler und Nenner des Ausdrucks auf der rechten Seite der Formel (2.7.3) durch h und geht dann mit $h \to 0$ zur Grenze über, so erhält man die Beziehung

$$F(y\,|\,x) = \lim_{h\to 0} P(Y < y\,|\,x \leqq X < x + h) = \frac{\int_{-\infty}^{y} f(x,y)\,dy}{f_1(x)}. \tag{2.7.5}$$

Der Ausdruck (2.7.5) ist für feste x-Werte *die bedingte Verteilungsfunktion* der Zufallsvariablen Y. Aus unseren Voraussetzungen folgt, daß sie eine Verteilungsfunktion einer stetigen Zufallsvariablen ist.

Ihre Dichtefunktion, die wir mit $f(y\,|\,x)$ bezeichnen wollen, wird durch die Formel

$$f(y\,|\,x) = \frac{f(x,y)}{f_1(x)} \tag{2.7.6}$$

gegeben.

Ebenso kann man die Verteilungsfunktion und die Dichtefunktion der bedingten Verteilung der Zufallsvariablen X unter der Bedingung $Y = y$ bestimmen.

Wir wollen (2.7.5) in der Gestalt

$$f_1(x)F(y\,|\,x) = \int_{-\infty}^{y} f(x,y)\,dy$$

schreiben. Daraus erhalten wir die Formel

$$\int_{-\infty}^{\infty} f_1(x)F(y\,|\,x)\,dx = F_2(y), \tag{2.7.7}$$

die eine Verallgemeinerung des Satzes 1.6.1 über die totale Wahrscheinlichkeit für stetige Zufallsvariable ist.

Wenn wir Zufallsvariable untersuchen, deren Dimension größer als 2 ist, können wir bedingte Verteilungen unter der Bedingung betrachten, daß eine gewisse Gruppe von Zufallsvariablen feste Werte annimmt. Als Beispiel wollen wir die dreidimensionale stetige Zufallsvariable (X_1, X_2, X_3) betrachten, deren Dichtefunktion $f(x_1, x_2, x_3)$ überall stetig und deren sämtliche Randdichtefunktionen stetig seien. In diesem Fall erhalten wir z. B. die folgenden bedingten Ver-

teilungsfunktionen:

$$F(x_3 \mid x_1, x_2) = \frac{\int\limits_{-\infty}^{x_3} f(x_1, x_2, x_3)\, dx_3}{\int\limits_{-\infty}^{\infty} f(x_1, x_2, x_3)\, dx_3}, \tag{2.7.8}$$

$$F(x_3, x_2 \mid x_1) = \frac{\int\limits_{-\infty}^{x_3} \int\limits_{-\infty}^{x_2} f(x_1, x_2, x_3)\, dx_2\, dx_3}{\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x_1, x_2, x_3)\, dx_2\, dx_3}. \tag{2.7.9}$$

Dabei nehmen wir selbstverständlich an, daß die Nenner auf der rechten Seite von (2.7.8) und (2.7.9) positiv sind. Analoge Formeln erhält man für die bedingten Dichtefunktionen.

Beispiel 2.7.1. Wir kehren zur Tabelle 2.6.1 zurück und wollen feststellen, wie die durch 3 teilbaren und durch 3 nichtteilbaren geraden Zahlen verteilt sind. Anders ausgedrückt fragen wir: Wie groß ist die Wahrscheinlichkeit, daß wir, wenn wir aufs Geratewohl ziehen, eine durch 3 teilbare Zahl erhalten, wenn wir schon vorher wissen, daß die Zahl gerade ist? Bei dieser Formulierung des Problems interessiert uns nicht die Verteilung aller 21 Zahlen, sondern nur die Verteilung eines gewissen Teiles, nämlich die Verteilung der 10 geraden Zahlen hinsichtlich ihrer Teilbarkeit durch 3.

Benutzen wir die Bezeichnungen aus Beispiel 2.6.1, so erhalten wir auf Grund von (2.7.2)

$$P(Y = 1 \mid X = 1) = \frac{P(Y = 1, X = 1)}{P(X = 1)} = \frac{p_{11}}{p_{1\cdot}} = \frac{3}{10},$$

$$P(Y = 0 \mid X = 1) = \frac{P(Y = 0, X = 1)}{P(X = 1)} = \frac{p_{12}}{p_{1\cdot}} = \frac{7}{10}.$$

Analog erhält man die bedingte Verteilung der ungeraden Zahlen in durch 3 teilbare und durch 3 nichtteilbare:

$$P(Y = 0 \mid X = 0) = \frac{P(Y = 0, X = 0)}{P(X = 0)} = \frac{p_{22}}{p_{2\cdot}} = \frac{7}{11},$$

$$P(Y = 1 \mid X = 0) = \frac{P(Y = 1, X = 0)}{P(X = 0)} = \frac{p_{21}}{p_{2\cdot}} = \frac{4}{11}.$$

Ebenso kann man die bedingten Verteilungen der durch 3 teilbaren sowie der durch 3 nichtteilbaren Zahlen in gerade und ungerade bestimmen.

C. Einen Spezialfall der bedingten Verteilung einer Zufallsvariablen X stellt die sogenannte gestutzte Verteilung von X dar.

Definition 2.7.1. Es sei X eine Zufallsvariable und S eine Borelsche Menge auf der x-Achse mit der Eigenschaft, daß $0 < P(X \in S) < 1$ gilt. Die für jedes reelle x durch den Ausdruck $P(X < x \mid X \in S)$ definierte bedingte Wahrscheinlichkeit nennen wir die *gestutzte Verteilung* der Zufallsvariablen X.

Ist X eine diskrete Zufallsvariable mit den Sprungstellen x_i und den Sprunghöhen p_i, dann hat die Wahrscheinlichkeitsfunktion der gestutzten Verteilung von X die Form

$$P(X = x_i | X \in S) = \frac{P(X = x_i, X \in S)}{P(X \in S)} = \begin{cases} \dfrac{p_i}{\sum\limits_{x_j \in S} p_j} & \text{für } x_i \in S, \\ 0 & \text{für } x_i \notin S. \end{cases}$$

Ist X eine stetige Zufallsvariable mit der Dichte $f(x)$, so gilt

$$P(X < x | X \in S) = \frac{P(X < x, X \in S)}{P(X \in S)} = \frac{\int\limits_{(-\infty, x) \cap S} f(x)\,dx}{\int\limits_S f(x)\,dx}.$$

Die Dichte $g(x)$ dieser Verteilung nimmt die folgende Form an:

$$g(x) = \begin{cases} \dfrac{f(x)}{\int\limits_S f(x)\,dx} & \text{für } x \in S, \\ 0 & \text{für } x \notin S. \end{cases}$$

Beispiel 2.7.2. Wir betrachten die Zufallsvariable X mit der Dichte

$$f(x) = \begin{cases} 1 \text{ für } 0 \leqq x \leqq 1, \\ 0 \text{ für } x < 0, \quad x > 1. \end{cases}$$

Für die Menge S wählen wir das Intervall $\left[0, \frac{1}{2}\right)$; dann gilt

$$P(X \in S) = \int\limits_0^{1/2} dx = \frac{1}{2}.$$

Die Dichte $g(x)$ der gestutzten Verteilung der Zufallsvariablen X hat die Form

$$g(x) = \begin{cases} 2 \text{ für } 0 \leqq x \leqq \dfrac{1}{2}, \\ 0 \text{ für } x < 0, x > \dfrac{1}{2}. \end{cases}$$

2.8. Unabhängige Zufallsvariable

A. In 1.7 haben wir den Begriff der Unabhängigkeit von zufälligen Ereignissen definiert. Wir übertragen diesen Begriff jetzt auf Zufallsvariable.

Mit $F(x,y), F_1(x)$ und $F_2(y)$ seien entsprechend die Verteilungsfunktionen der zweidimensionalen Zufallsvariablen (X, Y) und die Randverteilungsfunktionen der Zufallsvariablen X und Y bezeichnet.

Definition 2.8.1. Die Zufallsvariablen X und Y sind *unabhängig*, wenn für beliebige reelle Zahlenpaare (x,y) die Gleichung

$$F(x,y) = F_1(x)F_2(y) \tag{2.8.1}$$

gilt.

Es seien jetzt (a,b) und (c,d) mit $a < c$ und $b < d$ beliebige Punkte der x,y-Ebene. Es ist

$$P(a \leqq X < c) = F_1(c) - F_1(a),$$

$$P(b \leqq Y < d) = F_2(d) - F_2(b).$$

Multiplizieren wir die Seiten dieser Gleichungen miteinander und wenden wir die Formel (2.8.1) an, so erhalten wir

$$P(a \leqq X < c)P(b \leqq Y < d) = F(c,d) - F(c,b) - F(a,d) + F(a,b),$$

woraus unter Berücksichtigung von (2.5.3)

$$P(a \leqq X < c, b \leqq Y < d) = P(a \leqq X < c)P(b \leqq Y < d) \tag{2.8.2}$$

folgt.

Wir betrachten die zweidimensionale diskrete Zufallsvariable (X, Y) mit den Sprungstellen (x_i, y_k) und den Sprunghöhen p_{ik}; X und Y seien unabhängig. Betrachten wir die Formel (2.8.2) für den Spezialfall, daß sich die Rechtecke $(a \leqq X < c, b \leqq Y < d)$ auf einen Punkt $(X = a,\ Y = b)$ zusammenziehen, also für $c \to a$ und $d \to b$, und benutzen wir den Satz 1.3.4, so erhalten wir für beliebige Paare (x_i, y_k) die Gleichung

$$p_{ik} = P(X = x_i,\ Y = y_k) = P(X = x_i)P(Y = y_k) = p_{i\cdot}p_{\cdot k}. \tag{2.8.3}$$

Man zeigt leicht: Ist die Gleichung (2.8.3) für beliebige Zahlenpaare (x_i, y_k) erfüllt, so ist es auch die Gleichung (2.8.1).

Also haben wir folgendes Ergebnis erhalten: Ist (X, Y) eine diskrete Zufallsvariable, so ist für die Unabhängigkeit der Zufallsvariablen X und Y notwendig und hinreichend, daß die Gleichung (2.8.3) für beliebige Zahlenpaare (x_i, y_k) erfüllt ist.

Wenn wir (2.7.1) und (2.7.2) berücksichtigen, so erhalten wir für beliebige Zahlenpaare (i, k) die Beziehungen

$$\begin{aligned} P(X = x_i \mid Y = y_k) &= P(X = x_i), \\ P(Y = y_k \mid X = x_i) &= P(Y = y_k). \end{aligned} \tag{2.8.4}$$

Aus diesen Formeln folgt: Sind die Zufallsvariablen X und Y unabhängig, so ist die bedingte Verteilung der Zufallsvariablen X für alle Werte von Y dieselbe; also gibt uns ein Wert von Y keinerlei Informationen über die Verteilung der

Zufallsvariablen X, und umgekehrt ist die bedingte Verteilung von Y für alle Werte der Zufallsvariablen X dieselbe.

Ist die Zufallsvariable (X, Y) stetig, so erhalten wir durch Differentiation des Ausdrucks (2.8.1) nach x und y

$$\frac{\partial^2 F(x,y)}{\partial x\, \partial y} = f(x,y) = F_1'(x) F_2'(y) = f_1(x) f_2(y); \tag{2.8.5}$$

dabei schließen wir die Punktmenge aus, auf der die Dichtefunktion eventuell unstetig ist.

Es sei umgekehrt die Gleichung (2.8.5) erfüllt. Wir erhalten dann

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\, dx\, dy = \int_{-\infty}^{x} \int_{-\infty}^{y} f_1(x) f_2(y)\, dx\, dy$$

$$= \int_{-\infty}^{x} f_1(x)\, dx \int_{-\infty}^{y} f_2(y)\, dy = F_1(x)\, F_2(y)$$

und haben also das folgende Ergebnis bekommen: Ist (X, Y) eine stetige Zufallsvariable, deren Dichtefunktion $f(x,y)$ überall stetig ist, so ist für die Unabhängigkeit der Zufallsvariablen X und Y notwendig und hinreichend, daß die Gleichung (2.8.5) für alle (x,y) erfüllt ist.

Es seien S_1 bzw. S_2 beliebige Borelsche Mengen auf der x- bzw. y-Achse. Man kann beweisen, daß die Gleichung

$$P(X \in S_1, Y \in S_2) = P(X \in S_1) P(Y \in S_2) \tag{2.8.6}$$

gilt, wenn die Zufallsvariablen X und Y unabhängig sind.

Aus den Gleichungen (2.7.5) und (2.8.5) erhalten wir für jeden Wert von x

$$F(y|x) = F_2(y) \tag{2.8.7}$$

und analog für jedes y

$$F(x|y) = F_1(x). \tag{2.8.8}$$

Sind die Zufallsvariablen X und Y unabhängig, so hängt die bedingte Verteilungsfunktion der Zufallsvariablen Y unter der Bedingung $X = x$ nicht von x ab. Dasselbe gilt für die bedingte Verteilungsfunktion der Zufallsvariablen X. Auf Grund der Gleichungen (2.8.7) und (2.8.8) sowie der Gleichungen (2.8.4) erhält der Begriff der Unabhängigkeit eine sehr anschauliche Bedeutung.

Beispiel 2.8.1. Wir wollen zweimalige Würfe mit einer Münze betrachten. Die Zufallsvariable X soll den Wert 0 bzw. 1 annehmen, wenn als Ergebnis des ersten Wurfes Zahl bzw. Adler erscheint, die Zufallsvariable Y die Werte 0 bzw. 1, wenn beim zweiten Wurf Zahl bzw. Adler erscheint.

Als Ergebnis eines doppelten Wurfes kann die zweidimensionale Zufallsvariable (X, Y) einen der Werte

$$(1, 1), (1, 0), (0, 1), (0, 0)$$

annehmen.

Alle diese Ergebnisse treten mit derselben Wahrscheinlichkeit 1/4 auf. Jede der Zufallsvariablen X und Y kann einen der Werte 0 oder 1 mit der Wahrscheinlichkeit 1/2 annehmen. Wir erhalten also

$$P(X = 1, Y = 1) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(X = 1)\, P(Y = 1),$$

$$P(X = 1, Y = 0) = \frac{1}{4} = P(X = 1) P(Y = 0),$$

$$P(X = 0, Y = 1) = \frac{1}{4} = P(X = 0)\, P(Y = 1),$$

$$P(X = 0, Y = 0) = \frac{1}{4} = P(X = 0)\, P(Y = 0).$$

Die Zufallsvariablen X und Y sind also unabhängig.

Es seien X_1 und X_2 unabhängige Zufallsvariable. Wir betrachten die eindeutigen Funktionen $Y_1 = g_1(X_1)$ und $Y_2 = g_2(X_2)$. Dann sind Y_1 und Y_2 ebenfalls Zufallsvariable (vgl. die Fußnote auf S. 57), und wir wollen zeigen, daß Y_1 und Y_2 ebenfalls unabhängig sind.

Wir bezeichnen mit $g_1^{-1}(-\infty, y_1)$ und $g_2^{-1}(-\infty, y_2)$ die entsprechenden Borelschen Mengen, welche die Urbilder der Intervalle $(-\infty, y_1)$ und $(-\infty, y_2)$ bezüglich g_1 und g_2 sind. Dann erhalten wir

$$\begin{aligned} P(Y_1 < y_1, Y_2 < y_2) &= P\big(g_1(X_1) < y_1, g_2(X_2) < y_2\big) \\ &= P\big(X_1 \in g_1^{-1}(-\infty, y_1), X_2 \in g_2^{-1}(-\infty, y_2)\big) \\ &= P\big(X_1 \in g_1^{-1}(-\infty, y_1)\big) \cdot P\big(X_2 \in g_2^{-1}(-\infty, y_2)\big) \\ &= P(Y_1 < y_1) \cdot P(Y_1 < y_2). \end{aligned} \tag{2.8.9}$$

B. Der Begriff der Unabhängigkeit läßt sich auch auf endlich oder abzählbar viele zufällige Zufallsvariable ausdehnen.

Definition 2.8.2. Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ sind *unabhängig*, wenn für beliebige Punkte $(x_1, x_2, \ldots, x_n)$ die Gleichung

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1) F_2(x_2) \cdots F_n(x_n) \tag{2.8.10}$$

gilt; dabei ist $F(x_1, x_2, \ldots, x_n)$ die Verteilungsfunktion der Zufallsvariablen $(X_1, X_2, \ldots, X_n)$, und $F_1(x_1), \ldots, F_n(x_n)$ sind die Randverteilungsfunktionen der Zufallsvariablen $X_1, \ldots, X_n$.

Man kann die folgende Bemerkung machen: Sind die Zufallsvariablen X_1, $X_2, \ldots, X_n$ unabhängig, dann sind auch für jedes $s \leqq n$ die Zufallsvariablen $X_{k_1}, \ldots, X_{k_s} (1 \leqq k_1 < \cdots < k_s \leqq n)$ unabhängig. In der Tat folgt aus (2.8.10), wenn wir der einfacheren Schreibweise wegen $k_1 = 1$, $k_2 = 2, \ldots, k_s = s$ setzen, die Gleichung

$$F(x_1, x_2, \ldots, x_s, +\infty, \ldots, +\infty)$$
$$= F_1(x_1)F_2(x_2)\cdots F_s(x_s)F_{s+1}(+\infty)\cdots F_n(+\infty) = F_1(x_1)F_2(x_2)\cdots F_s(x_s).$$

Nun erklären wir die Unabhängigkeit für abzählbar viele Zufallsvariable.

Definition 2.8.3. Die Zufallsvariablen $X_1, X_2, X_3, \ldots$ sind *unabhängig*, wenn für jedes $n = 2, 3, 4, \ldots$ die Zufallsvariablen $X_1, X_2, \ldots, X_n$ unabhängig sind.

C. Der Begriff der Unabhängigkeit läßt sich folgendermaßen auf Zufallsvektoren übertragen:

Definition 2.8.4. Die Zufallsvektoren

$$X = (X_1, X_2, \ldots, X_j), \quad Y = (Y_1, Y_2, \ldots, Y_r)$$

sind *unabhängig*, wenn für alle $j + r$ reellen Zahlen $x_1, x_2, \ldots, x_j, y_1, y_2, \ldots, y_r$

$$F(x_1, x_2, \ldots, x_j, y_1, y_2, \ldots, y_r) = G(x_1, x_2, \ldots, x_j)H(y_1, y_2, \ldots, y_r)$$

gilt, wobei F, G bzw. H die Verteilungen der Zufallsvektoren $(X_1, X_2, \ldots, X_j,$ $Y_1, Y_2, \ldots, Y_r)$, $(X_1, X_2, \ldots, X_j)$ bzw. $(Y_1, Y_2, \ldots, Y_r)$ sind.

Es sei betont, daß aus der Unabhängigkeit der Zufallsvektoren $X = (X_1,$ $X_2, \ldots, X_j)$ und $Y = (Y_1, Y_2, \ldots, Y_r)$ keineswegs folgt, daß die Komponenten $X_1, X_2, \ldots, X_j$ des Vektors X oder die Komponenten $Y_1, Y_2, \ldots, Y_r$ des Vektors Y unabhängig sind.

Der Begriff der Unabhängigkeit läßt sich auch auf den Fall einer beliebigen endlichen oder abzählbaren Menge von Zufallsvektoren verallgemeinern. Die Definitionen sind den Definitionen 2.8.2 und 2.8.3 analog.

2.9. Funktionen von mehrdimensionalen Zufallsvariablen

A. Die Betrachtungen aus 2.4 lassen sich ohne Schwierigkeit auf Funktionen von mehrdimensionalen Zufallsvariablen übertragen. Wir wollen jetzt eine Formel für die zweidimensionale Dichte einer Funktion der stetigen Zufallsvariablen (X, Y) herleiten, die der Formel (2.4.19) entspricht, welche für eine Funktion einer eindimensionalen Zufallsvariablen galt.

Es sei

$$U_1 = g_1(X, Y), \quad U_2 = g_2(X, Y) \tag{2.9.1}$$

eine stetige und eineindeutige Transformation der Zufallsvariablen X und Y mit der zweidimensionalen Dichte $f(x, y)$. Wir wollen annehmen, daß die Funktionen g_1 und g_2 stetige partielle Ableitungen nach x und y besitzen. Ferner sei $[a \leqq X < b, c \leqq Y < d)$ ein Rechteck, auf dem die Funktionaldeterminate der Transformation (2.9.1) von Null verschieden ist. Wir bezeichnen mit $x = h_1(u_1, u_2)$ und $y = h_2(u_1, u_2)$ die Umkehrung der Transformation (2.9.1). Nach den Voraussetzungen sind die Funktionen h_1 und h_2 ebenfalls eindeutig und haben stetige partielle Ableitungen nach u_1 und u_2. Mit J bezeichnen wir die Funktionaldeterminante der Umkehrung der Transformation:

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u_1} & \dfrac{\partial x}{\partial u_2} \\ \dfrac{\partial y}{\partial u_1} & \dfrac{\partial y}{\partial u_2} \end{vmatrix}.$$

Die Funktionaldeterminante ist nach Voraussetzung endlich und stetig in dem Gebiet S der u_1, u_2-Ebene, in das das Rechteck $(a \leqq X < b, c \leqq Y < d)$ durch die Transformation (2.9.1) abgebildet wird. Wir erhalten weiter

$$\begin{aligned} P(a \leqq X < b, c \leqq Y < d) &= \int_a^b \int_c^d f(x, y)\, dx\, dy \\ &= \iint_{(S)} f[h_1(u_1, u_2), h_2(u_1, u_2)]\, |J|\, du_1\, du_2. \end{aligned} \tag{2.9.2}$$

Aus dieser Formel erkennen wir, daß die zweidimensionale Dichte der Variablen U_1 und U_2 die folgende Form hat:

$$h(u_1, u_2) = f[h_1(u_1, u_2), h_2(u_1, u_2)]\, |J|. \tag{2.9.3}$$

B. Wir beschäftigen uns jetzt mit den Verteilungen der *Summe*, der *Differenz*, des *Produktes* und des *Quotienten* zweier Zufallsvariabler. Diese Ausdrücke sind Beispiele für stetige Funktionen von mehrdimensionalen Zufallsvariablen, und außerdem sind sie äußerst wichtig für die Wahrscheinlichkeitsrechnung und deren Anwendungen.

Beispiel 2.9.1. Wir betrachten noch einmal das Würfeln mit einem Würfelpaar. Die Zufallsvariable X sei die Augenzahl des ersten Würfels und die Zufallsvariable Y die Augenzahl des zweiten Würfels. Jede der Veränderlichen X und Y kann die Werte $1, \ldots, 6$, jeden mit der Wahrscheinlichkeit $1/6$ annehmen.

Die zweidimensionale Veränderliche (X, Y) kann 36 Wertepaare (i, k) annehmen, wobei i und k alle Werte von 1 bis 6 durchlaufen. Wir bilden für jedes mögliche Wertepaar (i, k) die Summe $i + k$. Alle möglichen Werte der Summe $i + k$ bilden die Wertemenge einer neuen Zufallsvariablen, welche wir als *Summe* der Zufallsvariablen X und Y bezeichnen. Diese Summe ist eine eindimensionale Zufallsvariable. Sie kann die folgenden Werte annehmen:

$$2, 3, 4, \ldots, 11, 12.$$

Wir setzen $Z = X + Y$ und berechnen die Wahrscheinlichkeitsfunktion der Variablen Z. Wegen der Unabhängigkeit der Variablen X und Y (siehe 2.8) ist

$$P(Z = 2) = P(X = 1)\,P(Y = 1) = \frac{1}{36},$$

$$P(Z = 3) = P(X = 1,\, Y = 2) + P(X = 2,\, Y = 1) = \frac{1}{18},$$

$$P(Z = 4) = P(X = 1,\, Y = 3) + P(X = 2,\, Y = 2) + P(X = 3,\, Y = 1) = \frac{1}{12},$$

. . . .

$$P(Z = 12) = P(X = 6,\, Y = 6) = \frac{1}{36}.$$

Der Leser kann sich leicht davon überzeugen, daß tatsächlich

$$P(Z = 2) + P(Z = 3) + \cdots + P(Z = 12) = 1$$

ist.

Ebenso kann man beim Würfeln mit zwei Würfeln die Zufallsvariable $V = X - Y$ betrachten (X und Y haben die gleiche Bedeutung wie vorhin). Die Wertemenge der Zufallsvariablen V ist hier die Menge aller möglichen Werte der Differenz $i - k$. Die Zufallsvariable V kann die Werte

$$-5, -4, \ldots, 0, 1, 2, \ldots, 5$$

annehmen. Als Beispiel berechnen wir

$$\begin{aligned} P(V = 0) &= P(X = 1,\, Y = 1) + P(X = 2,\, Y = 2) + P(X = 3,\, Y = 3) \\ &\quad + P(X = 4,\, Y = 4) + P(X = 5,\, Y = 5) + P(X = 6,\, Y = 6) \\ &= \frac{1}{6}, \end{aligned}$$

$$P(V = -5) = P(X = 1,\, Y = 6) = \frac{1}{36}.$$

Wie der Leser leicht nachprüfen kann, gilt die Gleichung

$$\sum_{k=-5}^{5} P(V = k) = 1.$$

Ähnlich kann man die Wahrscheinlichkeitsfunktionen der Zufallsvariablen $T = X \cdot Y$ und $S = \frac{Y}{X}$ betrachten.

An dem obigen Beispiel erkennt man, was man unter der Summe, der Differenz, dem Produkt oder dem Quotienten zweier Zufallsvariabler zu verstehen hat.

Es ist also die Zufallsvariable $X + Y$ eine Funktion der zweidimensionalen Zufallsvariablen (X, Y). Zur Wertemenge der Zufallsvariablen $X + Y$ gehören alle möglichen Werte der Summe $x + y$ und nur diese, wobei x ein möglicher Wert der Zufallsvariablen X und y ein möglicher Wert der Zufallsvariablen Y ist. Ebenso kann man die Zufallsvariable definieren, die gleich der Summe endlich vieler Zufallsvariabler ist.

Die Wertemenge der Zufallsvariablen $X - Y$, die gleich der Differenz zweier Zufallsvariabler X und Y ist, besteht aus allen möglichen Werten von $x - y$, wobei x ein möglicher Wert von X und y ein möglicher Wert von Y ist. Analog definiert man das Produkt und den Quotienten zweier Zufallsvariabler.

C. Es sei die Verteilung der zweidimensionalen Zufallsvariablen (X, Y) bekannt. Wir wollen dann die Verteilungen derjenigen Zufallsvariablen bestimmen, die wir als Resultat der vier arithmetischen Operationen, ausgeführt an den Zufallsvariablen X und Y, erhalten.

Gesucht ist die Verteilungsfunktion der Summe

$$Z = X + Y \tag{2.9.4}$$

der Zufallsvariablen X und Y. Wir wollen also die Funktion

$$F(z) = P(X + Y < z)$$

bestimmen.

Wenn wir für einen gegebenen Wert von z in der x, y-Ebene die Gerade $x + y = z$ zeichnen (Abb. 2.9.1), ist $F(z)$ die Wahrscheinlichkeit dafür, daß der Punkt mit den Koordinaten (x, y) unterhalb dieser Geraden liegt.

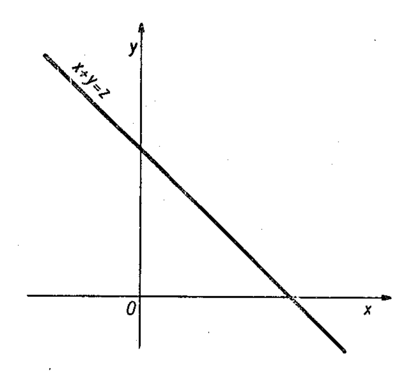


Abb. 2.9.1

Ist (X, Y) eine diskrete Zufallsvariable, die die Werte (x_i, y_k) annimmt, dann erhalten wir

$$F(z) = \sum_{x_i + y_k < z} P(X = x_i, Y = y_k), \tag{2.9.5}$$

wobei die Summation über alle Werte x_i, y_k zu erstrecken ist, die die Ungleichung $x_i + y_k < z$ erfüllen.

Es sei (X, Y) eine stetige Zufallsvariable, und $f(x, y)$, $f_1(x)$, $f_2(y)$ seien die Dichten der Zufallsvariablen (X, Y), X bzw. Y. Wir wollen die Gleichung (2.9.4) in der Form

$$X = X, \quad Z = X + Y \quad \text{oder} \quad X = X, \quad Y = Z - X \tag{2.9.6}$$

schreiben; dabei haben wir die Identität $X = X$ mit aufgeschrieben, um dieses Problem auf die schon behandelte Transformation (2.9.1) zurückzuführen. Es ist

$$J = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1 \quad \text{und} \quad |J| = 1.$$

Aus (2.9.3) folgt, daß die Dichte der Zufallsvariablen (X, Z) gleich

$$f(x, z - x) \tag{2.9.7}$$

ist. Um die Dichte $\psi(z)$ der Zufallsvariablen Z zu bestimmen, integrieren wir den Ausdruck (2.9.7) nach x von $-\infty$ bis ∞ und erhalten als Randdichte der zweidimensionalen Zufallsvariablen (X, Z)

$$\psi(z) = \int_{-\infty}^{\infty} f(x, z - x)\, dx. \tag{2.9.8}$$

Schließlich erhalten wir

$$F(z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f(x, z - x)\, dx\, dz. \tag{2.9.9}$$

Wenn die Zufallsvariablen X und Y unabhängig sind, ist

$$f(x, y) = f_1(x) f_2(y),$$

also

$$\psi(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z - x)\, dx \tag{2.9.8'}$$

und

$$F(z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_1(x) f_2(z - x)\, dx\, dz. \tag{2.9.9'}$$

In den Formeln (2.9.8), (2.9.8'), (2.9.9), (2.9.9') kann man wegen der Symmetrie der Summe x durch $z - y$ und $z - x$ durch y ersetzen.

Beispiel 2.9.2. Die Zufallsvariable X habe die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{2.9.10}$$

Statt des Ausdrucks $e^{-\frac{x}{2}}$ schreibt man auch

$$\exp\left(-\frac{1}{2}\,x^2\right).$$

In FICHTENHOLZ [1], S. 633, oder SMIRNOW [1], S. 219—223, findet der Leser einen Beweis für die Beziehung

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = 1.$$

Der Ausdruck (2.9.10) ist die Dichte der *Gaußschen Verteilung* oder auch *Normalverteilung*. Wir werden sie später im einzelnen untersuchen.

Die Zufallsvariable Y habe die Dichte

$$f(y) = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{y^2}{2}}.$$

Wir nehmen an, daß X und Y unabhängig sind; die Dichte von (X, Y) ist also gleich

$$f(x, y) = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}\,e^{-\frac{y^2}{2}} = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2+y^2}{2}}. \tag{2.9.11}$$

Wir betrachten die Zufallsvariable

$$Z = X + Y.$$

Nach Formel (2.9.8′) ist die Dichte der Zufallsvariablen Z

$$\psi(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(z-x)^2}{2}\right)dx$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(-\frac{2x^2 - 2zx + z^2}{2}\right)dx.$$

Wir benutzen die algebraische Umformung

$$2x^2 - 2zx + z^2 = \left(x\sqrt{2}\right)^2 - 2\left(x\sqrt{2}\right)\frac{z}{\sqrt{2}} + \left(\frac{z}{\sqrt{2}}\right)^2 + \frac{z^2}{2} = \left(x\sqrt{2} - \frac{z}{\sqrt{2}}\right)^2 + \frac{z^2}{2};$$

daraus ergibt sich

$$-\frac{2x^2 - 2zx + z^2}{2} = -\frac{\left(x\sqrt{2} - \frac{z}{\sqrt{2}}\right)^2}{2} - \frac{z^2}{4},$$

also

$$\psi(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left[-\frac{\left(x\sqrt{2} - \frac{z}{\sqrt{2}}\right)^2}{2}\right] \exp\left(-\frac{z^2}{4}\right) dx$$

$$= \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{z^2}{4}\right) \cdot \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{\left(x\sqrt{2} - \frac{z}{\sqrt{2}}\right)^2}{2}\right] dx.$$

Wenn wir in dem letzten Integral $u = x\sqrt{2} - \frac{z}{\sqrt{2}}$ substituieren, erhalten wir

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{\left(x\sqrt{2} - \frac{z}{\sqrt{2}}\right)^2}{2}\right] dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) du = 1,$$

woraus schließlich

$$\psi(z) = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{z^2}{4}\right)$$

folgt. Die Verteilungsfunktion $F(z)$ wird also durch die Formel

$$F(z) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{4}} dz$$

gegeben.

Wir überlassen es dem Leser zur Übung, die Formeln für die Verteilungsfunktion und für die Dichte der Differenz zweier Zufallsvariabler herzuleiten.

Wir betrachten jetzt das Produkt zweier Zufallsvariabler X und Y, die die zweidimensionale Dichte $f(x, y)$ und die Randdichten $f_1(x)$ bzw. $f_2(y)$ haben. Wir setzen

$$Z = XY \tag{2.9.12}$$

und schreiben diese Gleichung in Form eines Gleichungssystems:

$$X = X, \quad Y = \frac{Z}{X}.$$

Es ist

$$J = \begin{vmatrix} 0 & 1 \\ \frac{1}{x} & \frac{-z}{x^2} \end{vmatrix} = -\frac{1}{x},$$

also $|J| = \frac{1}{|x|}$.

Aus (2.9.3) folgt, daß die zweidimensionale Dichte der Zufallsvariablen (Z, X) gleich

$$f\left(x, \frac{z}{x}\right) \frac{1}{|x|} \tag{2.9.13}$$

ist. Die Dichte $\psi(z)$ der Zufallsvariablen Z erhalten wir durch Integration des Ausdrucks (2.9.13) nach x von $-\infty$ bis ∞:

$$\psi(z) = \int_{-\infty}^{\infty} f\left(x, \frac{z}{x}\right) \frac{1}{|x|} \, dx.. \tag{2.9.14}$$

Die Verteilungsfunktion von Z hat die Form

$$F(z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f\left(x, \frac{z}{x}\right) \frac{1}{|x|} \, dx \, dz. \tag{2.9.15}$$

Wenn X und Y unabhängig sind, dann nehmen die Gleichungen (2.9.14) bzw. (2.9.15) folgende Form an:

$$\psi(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_1(x) f_2\left(\frac{z}{x}\right) dx \tag{2.9.14'}$$

bzw.

$$F(z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} \frac{1}{|x|} f_1(x) f_2\left(\frac{z}{x}\right) dx \, dz. \tag{2.9.15'}$$

Wir können wegen der Symmetrie des Produkts in den Formeln (2.9.14), (2.9.14'), (2.9.15) und (2.9.15') x durch $\frac{z}{y}$, $\frac{z}{x}$ durch y und $|x|$ durch $|y|$ ersetzen.

Für den Quotienten zweier Zufallsvariabler X und Y

$$Z = \frac{X}{Y}$$

erhalten wir die Formeln

$$\psi(z) = \int_{-\infty}^{\infty} f(yz, y) \, |y| \, dy \tag{2.9.16}$$

und

$$F(z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f(yz, y) |y| \, dy \, dz. \tag{2.9.17}$$

Wenn die Zufallsvariablen X und Y unabhängig sind, ergibt sich

$$\psi(z) = \int_{-\infty}^{\infty} f_1(yz) f_2(y) |y| \, dy, \tag{2.9.16'}$$

$$F(z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_1(yz) f_2(y) |y| \, dy \, dz. \tag{2.9.17'}$$

Die Formeln für die Verteilung der Summe, der Differenz, des Produktes und des Quotienten zweier unabhängiger Zufallsvariabler X und Y ohne die Annahme, daß der Zufallsvektor (X, Y) stetig ist, sind in 6.5 angegeben. (Wegen der Beweise für diese Formeln siehe den Anhang.)

2.10. Schlußbemerkungen

In diesem Kapitel haben wir Zufallsvektoren betrachtet und damit Zufallsvariable, die Werte aus einem endlichdimensionalen euklidischen Raum annehmen. Kürzlich hat sich eine Theorie allgemeinerer Zufallsvariabler entwickelt, und zwar die Theorie der Zufallsvariablen, deren Werte in einem unendlichdimensionalen Raum, etwa in einem Funktionalraum liegen. Einige Bemerkungen zu diesem Thema werden wir in 8.1 behandeln. Die fundamentalen Informationen über diese Theorie finden wir in den Arbeiten von Kolmogoroff [8], Mourier [1], Prochorow [3], Getoor [2] sowie in den Monographien von Doob [5] und Loève [4].

2.11. Aufgaben und Ergänzungen

1. Es sei X eine Zufallsvariable. Ist dann auch $|X|$ eine Zufallsvariable?
2. Die Wahrscheinlichkeitsfunktion einer Zufallsvariablen X hat die Form

$$P(X = r) = e^{-\lambda} \frac{\lambda^r}{r!} \qquad (r = 0, 1, 2, \ldots).$$

Man bestimme die Wahrscheinlichkeitsfunktion der folgenden Zufallsvariablen: a) $Y = -X$, b) $Y = aX + b$, c) $Y = X^2$, d) $Y = \sqrt{X}$, e) $Y = X^l$ (l ganzzahlig).

3. Die Funktionen $F(x)$ bzw. $f(x)$ seien die Verteilung bzw. die Dichte der Zufallsvariablen X. Man bestimme die Verteilung und die Dichte der folgenden Zufallsvariablen: a) $Y = \frac{1}{X}$, b) $Y = e^X$, c) $Y = \sin X$, d) $Y = |X|$, e) $Y = F(X)$.
4. Die Verteilung des Vektors (X, Y) sei durch die folgenden Formeln gegeben:

$$P(X = 1, Y = 1) = P(X = 1, Y = 2) = P(X = 2, Y = 2) = \frac{1}{3}.$$

a) Man bestimme die Verteilungsfunktionen $F(x, y)$, $F_1(x)$ und $F_2(y)$.

b) Man stelle fest, ob die Punkte $\left(1, \frac{1}{2}\right)$, $(1, 3)$, $\left(2, \frac{1}{2}\right)$ und $(2, 3)$ Unstetigkeitspunkte von $F(x, y)$ sind.

5. Die Wahrscheinlichkeitsfunktion des Zufallsvektors (X, Y) hat die Form

$$P(X = 0, 1 \leqq Y \leqq 2) = P(1 \leqq X \leqq 2, Y = 0) = \frac{1}{2}.$$

a) Man bestimme die Unstetigkeitspunkte der Verteilungsfunktion $F(x, y)$.

b) Man überprüfe, ob das Rechteck mit den Eckpunkten $(0, 0)$, $(2, 0)$, $\left(0, \frac{1}{2}\right)$ und $\left(2, \frac{1}{2}\right)$ ein Stetigkeitsrechteck ist.

6. Man beweise die Formel (2.5.12)

a) für $n = 3$,

b) für ein beliebiges n.

7. Man gebe ein Beispiel dreier abhängiger Zufallsvariabler X, Y und Z an, die paarweise unabhängig sind.

Hinweis. Sollten dem Leser keine anderen Beispiele einfallen, so kann er das Beispiel 1.7.1 entsprechend modifizieren. Weitere Beispiele findet der Leser in der Arbeit von Geisser und Mantel [1].

8. Die Funktion $F(x, y)$ bzw. $f(x, y)$ seien die Verteilungsfunktion bzw. die Dichte eines Zufallsvektors (X, Y). Man bestimme die Verteilungsfunktion und die Dichte des Zufallsvektors (Z, U), der durch die folgenden Formeln definiert ist:

$$Z = aX + bY + c,$$
$$U = dX + eY + f.$$

9. Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien unabhängig. Man beweise:

a) Ist wenigstens eines der X_k stetig, so ist auch die Summe $X_1 + X_2 + \cdots + X_n$ stetig;

b) besitzt wenigstens eines der X_k eine stetige Verteilungsfunktion, so ist auch die Verteilungsfunktion der Summe $X_1 + X_2 + \cdots + X_n$ stetig.

Hinweis. Man verwende die Formel (6.5.6).

10. Man beweise, daß die Sätze a) und b) der vorhergehenden Aufgabe nicht mehr gelten, wenn man die Voraussetzung der Unabhängigkeit fallen läßt.

Hinweis. Man gebe ein Beispiel an.

3. DIE PARAMETER DER VERTEILUNG EINER ZUFALLSVARIABLEN

3.1. Der Mittelwert

Mit der Verteilung einer Zufallsvariablen hängen gewisse Größen, die Parameter der Verteilung, zusammen, die eine große Rolle in der mathematischen Statistik spielen. Solche Parameter sind die Momente und ihre Funktionen sowie die Lageparameter. Wir wollen sie der Reihe nach kennenlernen.

Es sei X eine Zufallsvariable. Wir betrachten die eindeutige Funktion $g(X)$ der Zufallsvariablen X.

Definition 3.1.1. Es sei X eine diskrete Zufallsvariable mit den Sprungstellen x_k und den Sprunghöhen p_k. Ist die Ungleichung

$$\sum_k p_k |g(x_k)| < \infty \tag{3.1.1'}$$

erfüllt, so nennen wir die Reihe

$$\sum_k p_k g(x_k) \tag{3.1.1}$$

den *Mittelwert* $E[g(X)]$ *der Zufallsvariablen* $g(X)$.

Definition 3.1.2. Es sei X eine stetige Zufallsvariable mit der Dichtefunktion $f(x)$. Die Funktion $g(x)$ sei im Riemannschen Sinne integrierbar. Ist die Ungleichung

$$\int_{-\infty}^{\infty} |g(x)| f(x)\, dx < \infty \tag{3.1.2'}$$

erfüllt, so nennen wir das Integral

$$\int_{-\infty}^{\infty} g(x) f(x)\, dx \tag{3.1.2}$$

den *Mittelwert* $E[g(X)]$ *der Zufallsvariablen* $g(X)$.

Wir bemerken noch einmal, daß der Mittelwert der Zufallsvariablen $g(X)$ nicht existiert, wenn zwar (3.1.1) existiert, die Ungleichung (3.1.1′) aber nicht erfüllt ist.

Ebenso existiert auch $E[g(X)]$ nicht, wenn zwar (3.1.2) existiert (konvergiert), die Ungleichung (3.1.2′) aber nicht erfüllt ist.

Es sei $g(X) = X$. Die durch (3.1.1) bzw. (3.1.2) definierte Größe nennen wir den *Mittelwert der Zufallsvariablen* X.

Den Mittelwert $E(X)$ nennt man manchmal die *mathematische Erwartung*[1]) oder auch den *Erwartungswert der Zufallsvariablen* X. (Wegen der Definition von $E(X)$ siehe auch den Anhang.)

Wir zeigen, daß die Definition des Mittelwertes der Funktion $g(X)$ eindeutig ist: Setzen wir $Y = g(X)$ und berechnen wir den Mittelwert $E(Y)$ unmittelbar aus der Verteilung der Zufallsvariablen Y, so erhalten wir dasselbe Ergebnis, nämlich $E[g(X)]$, als wenn wir diesen Mittelwert nach der Definition 3.1.1 bzw. 3.1.2 aus der Verteilung von X berechnen. Wir beschränken uns auf den Beweis für den Fall einer diskreten Zufallsvariablen.

Es habe X also die Sprungstellen x_k und die Sprunghöhen p_k und Y die Sprungstellen y_i und die Sprunghöhen q_j. Dann ist $q_j = \sum_{g(x_k)=y_j} p_k \; (j = 1, 2, \ldots)$; dabei betrifft die Summationsvorschrift alle k, für die die Gleichung $g(x_k) = y_j$ gilt. Da nach der Voraussetzung über die Existenz des Mittelwertes $E[g(X)]$ die Reihe $\sum_k p_k g(x_k)$ absolut konvergiert, kann man ihre Glieder beliebig umordnen. Also erhalten wir

$$E(Y) = \sum_j q_j y_j = \sum_k p_k g(x_k) = E[g(X)].$$

Wir bringen einige Beispiele einer Mittelwertberechnung.

Beispiel 3.1.1. Wir nehmen an, daß die Zufallsvariable X zwei Werte annehmen kann: $x_1 = -1$ mit der Wahrscheinlichkeit $p_1 = 0{,}1$ und $x_2 = +1$ mit der Wahrscheinlichkeit $p_2 = 0{,}9$.

Der Mittelwert dieser Zufallsvariablen X ist

$$E(X) = 0{,}1 \cdot (-1) + 0{,}9 \cdot (+1) = 0{,}8.$$

Beispiel 3.1.2. Die Zufallsvariable X kann die Werte $k = 0, 1, 2, \ldots$ mit der Wahrscheinlichkeit

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{3.1.3}$$

annehmen, wobei λ eine gewisse positive Konstante ist. Die Formel (3.1.3) stellt eine sogenannte *Poissonsche Verteilung* dar.

Wir berechnen den Mittelwert der Zufallsvariablen X:

$$E(X) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \tag{3.1.4}$$

Beispiel 3.1.3. Die Zufallsvariable X kann die Werte $r = 0, 1, 2, \ldots, n$ mit der Wahrscheinlichkeit

$$P(X = r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \tag{3.1.5}$$

[1]) Das Symbol $E(X)$ rührt vom französischen „espérance“ (Hoffnung) her.

annehmen; dabei ist p eine gewisse Konstante, die die Ungleichung $0 < p < 1$ erfüllt. Eine derartige Verteilung nennen wir *Binomialverteilung*, da auf der rechten Seite der Gleichung (3.1.5) die Glieder des nach dem binomischen Satz entwickelten Binoms $[p + (1 - p)]^n$ stehen.

In einer Warenladung befinden sich gute und fehlerhafte Exemplare. Die Wahrscheinlichkeit, daß ein willkürlich gewähltes Exemplar gut ist, betrage p. Wir setzen voraus, daß bei einer zufälligen Auswahl von n Exemplaren die aufeinanderfolgenden Ereignisse, die darin bestehen, daß man beim k-ten Mal ($k \leqq n$) ein gutes Exemplar zieht, voneinander unabhängig sind (dies kann man zum Beispiel dadurch erreichen, daß man das schon gewählte Exemplar vor der Wahl des nächsten zurückgibt). Die Wahrscheinlichkeit, daß unter den n gewählten Exemplaren r gute vorhanden sind, wird durch die Formel (3.1.5) angegeben. Die Herleitung dieser Formel ist sehr einfach.

Wir bezeichnen ein fehlerhaftes Exemplar mit f und ein gutes mit g. Da die einzelnen aufeinanderfolgenden Ereignisse f und g unabhängig sein sollen, erhalten wir für die Wahrscheinlichkeit einer bestimmten Buchstabenanordnung, in der r-mal der Buchstabe g und $(n - r)$-mal der Buchstabe f auftritt, die Formel

$$p^r (1 - p)^{n-r}.$$

Die Gesamtzahl solcher Anordnungen (r-mal g und $(n - r)$-mal f) ist gleich

$$\binom{n}{r} = \frac{n!}{r!(n - r)!}.$$

Wir erhalten also (je zwei verschiedene Anordnungen schließen einander natürlich aus) die Formel (3.1.5).

Der Mittelwert dieser Zufallsvariablen X ist gleich

$$\begin{aligned} E(X) &= \sum_{r=0}^{n} r \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \\ &= \sum_{r=1}^{n} r \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} = np \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!\,(n-r)!} p^{r-1} (1-p)^{n-r} \\ &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k} = np[p + (1-p)]^{n-1} = np. \end{aligned} \tag{3.1.6}$$

Beispiel 3.1.4. Die Zufallsvariable X sei stetig und habe die Dichte $f(x)$, die durch die Formel

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

gegeben wird. Der Mittelwert dieser Zufallsvariablen ist gleich

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-x) e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \left[e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} = 0.$$

Wir geben jetzt zwei Beispiele für Zufallsvariable an, die keinen Mittelwert haben.

Beispiel 3.1.5. Die Zufallsvariable X möge die Werte $x_k = (-1)^k \frac{2^k}{k}$ $(k = 1, 2, \ldots)$ mit den Wahrscheinlichkeiten $p_k = \frac{1}{2^k}$ annehmen. Wir erhalten

$$\sum_{k=1}^{\infty} p_k x_k = \sum_{k=1}^{\infty} (-1)^k \frac{1}{k} = -\log 2.$$

Für diese Zufallsvariable existiert $E(X)$ jedoch nicht; denn es ist

$$\sum_{k=1}^{\infty} p_k |x_k| = \sum_{k=1}^{\infty} \frac{1}{k} = \infty. \tag{3.1.7}$$

Beispiel 3.1.6. Wir betrachten die Zufallsvariable $Y = |X|$, wobei X dieselbe Verteilung wie im vorhergehenden Beispiel hat. Aus (3.1.7) erkennt man, daß Y keinen Mittelwert hat.

In Kapitel 5 bringen wir ein Beispiel für eine stetige Zufallsvariable, die keinen Mittelwert besitzt.

3.2. Die Momente

A. Der in 3.1 behandelte Mittelwert $E(X)$ gehört zu einer Gruppe von Parametern, die *Momente* genannt werden.

Definition 3.2.1. Wir nennen den Mittelwert der Funktion $g(X) = X^k$ das *Moment k-ter Ordnung* der Zufallsvariablen X und schreiben $m_k = E(X^k)$.

Ist also X eine diskrete Zufallsvariable mit den Sprungstellen x_l und den Sprunghöhen p_l, dann ist nach (3.1.1) das Moment k-ter Ordnung gleich

$$m_k = E(X^k) = \sum_l x_l^k p_l. \tag{3.2.1}$$

Ist die Zufallsvariable X stetig mit der Dichte $f(x)$, dann ist nach (3.1.2) das k-te Moment gleich

$$m_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x)\,dx. \tag{3.2.2}$$

Wir erinnern daran, daß für die Existenz des Moments m_k die absolute Konvergenz der Reihe (3.2.1) oder des Integrals (3.2.2) notwendig ist. Daraus folgt: Existiert das Moment k-ter Ordnung, so existieren auch alle Momente einer Ordnung kleiner als k.

Wie wir sehen, besteht für den Fall, daß das Moment k-ter Ordnung der Zufallsvariablen X existiert, die Relation

$$\lim_{a\to\infty} a^k P(|X| > a) = 0 \quad (a > 0).$$

Es gilt nämlich, falls wir die Existenz des Moments k-ter Ordnung berücksichtigen (den Beweis bringen wir für stetige Zufallsvariable mit der Dichte $f(x)$),

$$\lim_{a\to\infty} a^k P(|X| > a) \leqq \lim_{a\to\infty} \int_{|x|>a} |x|^k f(x)\,dx = 0.$$

Die bewiesene Relation können wir in der anschaulicheren Form

$$P(|X| > a) = o\left(\frac{1}{a^k}\right)$$

schreiben. Wie man sieht, hängt die Existenz der Momente von der Wahrscheinlichkeit ab, mit welcher die Zufallsvariable, absolut genommen, große Werte annimmt. Ist die Menge der möglichen Werte, die die Zufallsvariable annimmt, beiderseitig beschränkt, so existieren selbstverständlich Momente beliebiger Ordnung.

Es seien $g_1(X)$ und $g_2(X)$ zwei eindeutige Funktionen der Zufallsvariablen X, und es mögen die Mittelwerte $E[g_1(X)]$ und $E[g_2(X)]$ existieren. Wir werden zeigen, daß die Gleichung

$$E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)] \tag{3.2.3}$$

besteht.

Zum Beweis beschränken wir uns auf den Fall einer stetigen Zufallsvariablen X. Dann ist

$$E[g_1(X) + g_2(X)] = \int_{-\infty}^{\infty} [g_1(x) + g_2(x)] f(x)\, dx. \tag{3.2.4}$$

Da nach Voraussetzung die beiden Ungleichungen

$$\int_{-\infty}^{\infty} |g_1(x)|\, f(x)\, dx < \infty,$$

$$\int_{-\infty}^{\infty} |g_2(x)|\, f(x)\, dx < \infty$$

erfüllt sind, existiert auch der Mittelwert der Summe $g_1(X) + g_2(X)$; denn es ist

$$\int_{-\infty}^{\infty} |g_1(x) + g_2(x)|\, f(x)\, dx \leqq \int_{-\infty}^{\infty} |g_1(x)|\, f(x)\, dx + \int_{-\infty}^{\infty} |g_2(x)|\, f(x)\, dx.$$

Also erhalten wir aus (3.2.4)

$$E[g_1(X) + g_2(X)] = \int_{-\infty}^{\infty} g_1(x) f(x)\, dx + \int_{-\infty}^{\infty} g_2(x) f(x)\, dx$$

$$= E[g_1(X)] + E[g_2(X)].$$

Die Formel (3.2.3) ist damit bewiesen. Man kann sie leicht auf endlich viele Summanden verallgemeinern.

Es wird dem Leser ohne Schwierigkeiten gelingen, die Formel

$$E[(aX)^k] = a^k E(X^k), \tag{3.2.5}$$

wobei a ein konstanter Koeffizient ist, zu beweisen.

Wenn wir außerdem noch berücksichtigen, daß für eine beliebige Konstante b die Gleichung

$$E(b) = b$$

gilt, so erhalten wir

$$E(aX + b) = aE(X) + b. \tag{3.2.6}$$

Beispiel 3.2.1. Die Zufallsvariable X kann die Werte 2 und 4 mit den Wahrscheinlichkeiten $P(X = 2) = 0{,}2$ und $P(X = 4) = 0{,}8$ annehmen. Gesucht ist $E(X^2)$. Aus (3.2.1) erhalten wir

$$E(X^2) = 4 \cdot 0{,}2 + 16 \cdot 0{,}8 = 13{,}6.$$

Beispiel 3.2.2. Die Zufallsvariable X sei normal verteilt mit der Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Gesucht ist $E(X^2)$. Benutzen wir die Formel (3.2.2) und integrieren partiell, so erhalten wir

$$E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-x) \left(-x e^{-\frac{x^2}{2}}\right) dx$$

$$= \frac{1}{\sqrt{2\pi}} \left[-x e^{-\frac{x^2}{2}}\right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1.$$

Man kann auch Ausdrücke der Form

$$E[(X - c)^k] \tag{3.2.7}$$

berechnen, wobei c eine beliebige Konstante ist.

Definition 3.2.2. Den Ausdruck (3.2.7) nennen wir *Moment k-ter Ordnung in bezug auf den Punkt c.*

Eine besonders wichtige Rolle spielen die Momente in bezug auf den Mittelwert, das heißt in bezug auf den Punkt $c = m_1 = E(X)$.

Definition 3.2.3. Die Momente in bezug auf den Mittelwert heißen *zentrale Momente.*

Wir bezeichnen sie mit μ_k. Es ist also

$$\mu_k = E[X - E(X)]^k. \tag{3.2.8}$$

Die Momente in bezug auf den Punkt $c = 0$ nennen wir *gewöhnliche* Momente.

Wenn wir den Ausdruck (3.2.8) umformen, können wir nach Berücksichtigung der Gleichung (3.2.3) die zentralen Momente durch ihre gewöhnlichen Momente ausdrücken. Für die ersten drei zentralen Momente erhalten wir

$$\begin{aligned}\mu_1 &= E[X - E(X)] = E(X - m_1) = E(X) - m_1 = m_1 - m_1 = 0,\\ \mu_2 &= E[X - E(X)]^2 = E[(X - m_1)^2]\\ &= E(X^2) - 2m_1E(X) + m_1^2 = m_2 - m_1^2,\\ \mu_3 &= E[X - E(X)]^3 = E[(X - m_1)^3]\\ &= E(X^3) - 3m_1E(X^2) + 3m_1^2E(X) - m_1^3\\ &= m_3 - 3m_1m_2 + 3m_1^3 - m_1^3 = m_3 - 3m_1m_2 + 2m_1^3.\end{aligned} \tag{3.2.9}$$

Wir wollen uns eingehender mit dem Fall $k = 2$ beschäftigen. Das Moment

$$E[(X - c)^2]$$

nennen wir auch die *mittlere quadratische Abweichung* der Zufallsvariablen X vom Punkt c.

Definition 3.2.4. Das zentrale Moment zweiter Ordnung oder die mittlere quadratische Abweichung der Zufallsvariablen X vom Mittelwert m nennen wir *Dispersion* oder *Varianz.*

Es ist üblich, die Dispersion mit $D^2(X)$ oder mit σ^2 zu bezeichnen. Auch die Bezeichnung $\mathbf{D}(X)$ findet sich in der Literatur. Aus den Formeln (3.2.9) erhalten wir

$$D^2(X) = \sigma^2 = m_2 - m_1^2. \tag{3.2.10}$$

Der Parameter σ^2 ist ein Maß für die Streuung einer Zufallsvariablen um ihren Mittelwert. Je mehr sich eine Verteilung um den Mittelwert häuft, desto kleiner ist der Wert von σ^2.

Definition 3.2.5. Die Quadratwurzel aus der Dispersion nennen wir die *Standardabweichung.*

Beispiel 3.2.3. Wir berechnen die Dispersion für eine Binomialverteilung (siehe Beispiel 3.1.3), für die

$$P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r} \qquad (r = 0, 1, \ldots, n)$$

ist. Gehen wir ganz analog wie bei der Berechnung von $E(X)$ in Beispiel 3.1.3 vor, so erhalten wir

$$E(X^2) = \sum_{r=0}^{n} r^2 \binom{n}{r} p^r (1 - p)^{n-r} = pn(q + pn),$$

wobei $q = 1 - p$ ist. Es ist $E(X) = pn$. Auf Grund von (3.2.10) erhalten wir daher die Formel

$$D^2(X) = pn(q + pn) - p^2n^2 = pqn. \tag{3.2.11}$$

Es sei etwa $n = 2$. Die Zufallsvariable kann also einen der Werte 0, 1, 2 annehmen. Ist außerdem $p = \frac{1}{2}$, so folgt $E(X) = \frac{1}{2} \cdot 2 = 1$.

Die durch die Größen $n = 2$, $p = \frac{1}{2}$ gegebene Verteilung ist in Abb. 3.2.1 dargestellt; dabei sind die Werte der Zufallsvariablen als Abszissen und die Wahrscheinlichkeiten dieser Werte als Ordinaten gewählt worden.

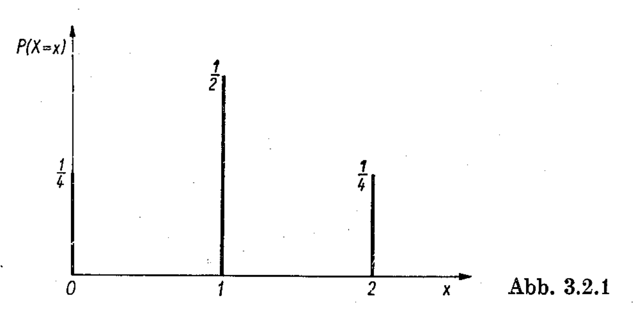


Abb. 3.2.1

Die Dispersion der Verteilung beträgt auf Grund der Gleichung (3.2.11)

$$D^2(X) = \frac{1}{2}.$$

Es sei nun $n = 3$ und $p = \frac{1}{3}$. Die Zufallsvariable X kann dann die Werte 0, 1, 2, 3 annehmen. Wir erhalten $E(X) = \frac{1}{3} \cdot 3 = 1$. Der Mittelwert ist derselbe wie vorher; die Dispersion ist aber größer; sie beträgt

$$D^2(X) = \frac{1}{3} \frac{2}{3} \cdot 3 = \frac{2}{3}.$$

Die entsprechende Verteilung ist in Abb. 3.2.2 dargestellt.

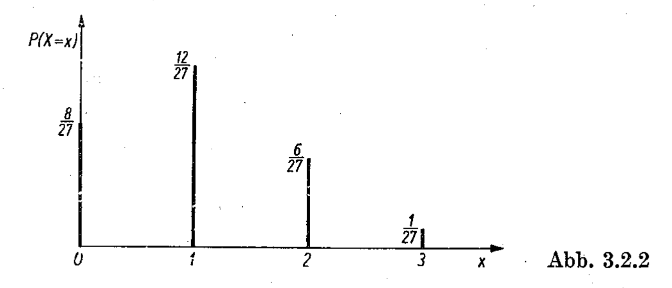


Abb. 3.2.2

Wenn wir Abb. 3.2.1 und 3.2.2 vergleichen, sehen wir, daß im zweiten Fall die Streuung der Werte der Zufallsvariablen X um den Mittelwert 1 größer ist als im ersten Fall. Durch diese Beispiele wird erläutert, daß die Dispersion als Streuungsmaß dienen kann.

Die Dispersion besitzt die folgende wichtige Eigenschaft: Für jedes $c \neq m_1$ gilt die Ungleichung

$$D^2(X) < E[(X - c)^2]. \tag{3.2.12}$$

Aus (3.2.3), (3.2.5) und (3.2.9) folgt nämlich

$$\begin{aligned} E[(X - c)^2] &= E[(X - m_1 + m_1 - c)^2] \\ &= E[(X - m_1)^2] + 2(m_1 - c)E(X - m_1) + (m_1 - c)^2 \\ &= D^2(X) + (m_1 - c)^2. \end{aligned}$$

Wenn wir berücksichtigen, daß nach Voraussetzung $(m_1 - c)^2 \neq 0$ ist, so erhalten wir die Ungleichung (3.2.12).

Wir wollen jetzt die Dispersion einer linearen Funktion der Zufallsvariablen X berechnen. Es sei $Y = X + c$. Dann erhalten wir

$$\begin{aligned} D^2(Y) &= E[Y - E(Y)]^2 \\ &= E[X + c - E(X) - E(c)]^2 \\ &= E[X - E(X)]^2 = D^2(X). \end{aligned} \tag{3.2.13}$$

Wir sagen, die Dispersion sei *translationsinvariant*, da eine Verschiebung ihren Wert nicht ändert.

Es sei nun

$$Y = aX + b.$$

Berücksichtigen wir (3.2.3), (3.2.5) und (3.2.10), so erhalten wir

$$\begin{aligned} D^2(Y) &= E[(aX + b)^2] - [E(aX + b)]^2 = a^2 E(X^2) - a^2[E(X)]^2 \\ &= a^2\left[E(X^2) - \left(E(X)\right)^2\right] = a^2 D^2(X). \end{aligned} \tag{3.2.14}$$

Definition 3.2.6. Eine Zufallsvariable X, für die die Gleichungen

$$E(X) = 0, \quad D^2(X) = 1$$

gelten, nennt man *normiert* oder *standardisiert*.

Hat eine Zufallsvariable X den Mittelwert m_1 und die Standardabweichung σ, dann ist die durch

$$Y = \frac{X - m_1}{\sigma}$$

definierte Zufallsvariable Y normiert. Aus den Formeln (3.2.6) und (3.2.14) erhalten wir nämlich $E(Y) = 0$ und $D^2(Y) = 1$.

Als Streuungsmaß kann außer der Standardabweichung die sogenannte *mittlere Abweichung* dienen, die durch

$$\sum_{-\infty<x_i<\infty} p_i |x_i - m_1| \tag{3.2.15}$$

definiert wird, wenn X eine diskrete Zufallsvariable mit den Sprungstellen x_i ist, bzw. durch

$$\int_{-\infty}^{\infty} |x - m_1| f(x)\, dx, \tag{3.2.15'}$$

wenn X eine stetige Zufallsvariable ist.

Im zweiten Teil dieses Buches werden wir uns mit Kriterien befassen, die zu entscheiden erlauben, welches dieser Streuungsmaße besser ist.

Manchmal ist es vorteilhafter, die Streuung nicht unmittelbar durch die Standardabweichung, sondern durch das Verhältnis dieser Abweichung zum Mittelwert zu charakterisieren. Dieses Verhältnis ist ebenfalls ein Verteilungsparameter.

Definition 3.2.7. Das Verhältnis der Standardabweichung zum Mittelwert nennen wir den *Variationskoeffizienten* und bezeichnen ihn mit v. Es ist also

$$v = \frac{\sigma}{m_1}. \tag{3.2.16}$$

Der Variationskoeffizient v ist gleich der Standardabweichung, falls der Mittelwert gleich Eins ist. Mit anderen Worten: Der Variationskoeffizient ist ein Streuungsmaß mit dem Mittelwert als Einheit.

Beispiel 3.2.4. Wir haben vorhin festgestellt, daß für die Binomialverteilung der Mittelwert gleich np (siehe Beispiel 3.1.3), die Standardabweichung dagegen gleich $\sqrt{nqp}$ ist (siehe Beispiel 3.2.3). Der Variationskoeffizient wird also für die Binomialverteilung durch die Formel

$$v = \frac{\sqrt{npq}}{np} = \sqrt{\frac{q}{np}}$$

gegeben.

B. Wir führen jetzt den Begriff einer symmetrischen Verteilung ein.

Definition 3.2.8. Wir sagen, daß die Zufallsvariable X *symmetrisch* verteilt ist, wenn ein Punkt a existiert, so daß die Verteilungsfunktion $F(x)$ von X für jeden Wert von x die Gleichung

$$F(a - x) = 1 - F(a + x) - P(X = a + x)$$

erfüllt. Den Punkt a nennen wir *Symmetriezentrum*.

Insbesondere gilt, falls $a = 0$, für jeden Wert von x

$$F(-x) = 1 - F(x) - P(X = x).$$

Wenn eine stetige Zufallsvariable symmetrisch verteilt ist, erfüllt ihre Dichte $f(x)$ die Gleichung (die eventuellen Unstetigkeitspunkte von $f(x)$ ausgenommen)

$$f(a - x) = f(a + x).$$

Ist eine diskrete Zufallsvariable X symmetrisch verteilt, so liegen ihre Sprungstellen und Sprunghöhen symmetrisch in bezug auf das Symmetriezentrum. Also können wir abschließend zusammenfassen:

Ist eine Zufallsvariable symmetrisch verteilt und besitzt sie einen endlichen Mittelwert, so ist dieser Mittelwert auch gleichzeitig Symmetriezentrum.

Daraus folgt weiter: Für eine symmetrische Verteilung sind die zentralen Momente ungerader Ordnung, wenn sie existieren, gleich Null.

Manchmal ist es nötig, den Grad der Asymmetrie einer Verteilung abzuschätzen. Da für eine symmetrische Verteilung alle ungeraden zentralen Momente gleich Null sind, kann also die Größe eines jeden ungeraden Moments als Maß für die Asymmetrie dienen. Es ist vorteilhaft, diese Größe mit einem Maßstab zu messen, in dem die Einheit gleich der Standardabweichung ist.

Deshalb nimmt man gewöhnlich als Maß für die Asymmetrie den durch

$$\gamma = \frac{\mu_3}{\sigma^3} \tag{3.2.17}$$

definierten Ausdruck und nennt ihn den *Koeffizienten der Asymmetrie* oder kurz die *Schiefe*. Dieser Koeffizient kann positiv oder negativ sein; je nach dem Vorzeichen sprechen wir also von einer positiven oder negativen Schiefe.

C. Wir machen nun einige Bemerkungen zum Problem, die Verteilungsfunktion $F(x)$ einer Zufallsvariablen X eindeutig zu bestimmen, wenn alle Momente $m_k = E(X^k)$ $(k = 1, 2, \ldots)$ existieren und bekannt sind. Wir begnügen uns hier damit, den nachstehenden Satz ohne Beweis anzugeben (siehe Aufgabe 4.8.10):

Satz 3.2.1. *Es mögen die Momente* m_k $(k = 1, 2, \ldots)$ *der Zufallsvariablen* X *existieren, und es sei die Reihe*

$$\sum_{k=1}^{\infty} \frac{m_k}{k!} r^k$$

für ein gewisses $r > 0$ *absolut konvergent. Dann bestimmt die Menge* $\{m_k\}$ *der Momente die Verteilungsfunktion* $F(x)$ *der Zufallsvariablen* X *eindeutig.*

Insbesondere folgt aus diesem Satz: Gilt für eine gewisse Konstante M die Ungleichung $|m_k| \leqq M^k$ $(k = 1, 2, \ldots)$, so bestimmen die Momente die Verteilungsfunktion $F(x)$ eindeutig. Ist somit die Menge aller möglichen Werte der Zufallsvariablen X nach beiden Seiten beschränkt, so ist $F(x)$ durch $\{m_k\}$ eindeutig bestimmt.

Existieren jedoch nicht alle diese Momente, so genügen die existierenden Momente nicht, um die Verteilungsfunktion $F(x)$ zu bestimmen, wie dies aus dem nachstehenden Beispiel hervorgeht.

Beispiel 3.2.5. Die Zufallsvariable X möge die Werte $x_k = 2^k/k^2$ $(k = 1, 2, \ldots)$ mit der Wahrscheinlichkeit $p_k = 1/2^k$ annehmen.

Es gilt

$$E(X) = \sum_{k=1}^{\infty} x_k p_k = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty,$$

$$\sum_{k=1}^{\infty} x_k^2 p_k = \sum_{k=1}^{\infty} \frac{2^k}{k^4} = \infty.$$

Nun möge die Zufallsvariable Y die Werte $y_k = 2^{k+1}/k^2$ $(k = 1, 2, \ldots)$ mit der Wahrscheinlichkeit $p'_k = 1/2^{k+1}$ und den Wert 0 mit der Wahrscheinlichkeit $\frac{1}{2}$ annehmen. Dann gilt

$$E(Y) = \sum_{k=1}^{\infty} y_k p'_k = \sum_{k=1}^{\infty} \frac{1}{k^2} = E(X),$$

$$\sum_{k=1}^{\infty} y_k^2 p'_k = \sum_{k=1}^{\infty} \frac{2^{k+1}}{k^4} = \infty.$$

Die Zufallsvariablen X und Y haben gleiche Momente erster Ordnung und weisen keine Momente von höherer als erster Ordnung auf. Die Verteilungen dieser Variablen sind aber offenbar voneinander verschieden.

Dem Problem, eine Verteilung mit Hilfe der Momente zu bestimmen, sind viele mathematische Arbeiten gewidmet. Es seien hier die Arbeiten von Stieltjes [1] und Hamburger [1] erwähnt.

Eine Bedingung dafür, daß die Verteilung einer zweidimensionalen Zufallsvariablen durch die Momente bestimmt ist, haben Cramér und Wold [1] angegeben.

3.3. Die Tschebyscheffsche Ungleichung

Die Rolle der Standardabweichung als Parameter, der die Streuung charakterisiert, wird besonders deutlich an der berühmten *Tschebyscheffschen Ungleichung*

$$P(|X - m_1| \geqq k\sigma) \leqq \frac{1}{k^2}, \tag{3.3.1}$$

die für jedes positive k gilt. Die Tschebyscheffsche Ungleichung ist eine Folgerung aus dem nachstehenden Satz:

Satz 3.3.1. *Nimmt eine Zufallsvariable Y nur nichtnegative Werte an und besitzt sie einen endlichen Mittelwert $E(Y)$, so ist für jede positive Zahl K die Ungleichung*

$$P(Y \geqq K) \leqq \frac{E(Y)}{K} \tag{3.3.2}$$

erfüllt.

Beweis. Wir beweisen diesen Satz für stetige Zufallsvariable. (Der Beweis für diskrete Zufallsvariable verläuft analog.) Es ist

$$E(Y) = \int_0^\infty y f(y)\,dy = \int_0^K y f(y)\,dy + \int_K^\infty y f(y)\,dy$$

$$\geqq \int_K^\infty y f(y)\,dy \geqq K \int_K^\infty f(y)\,dy = K P(Y \geqq K).$$

Daraus folgt sofort die Ungleichung (3.2.2).

Es sei nun X eine Zufallsvariable mit dem Mittelwert $E(X) = m_1$ und der Standardabweichung σ. Wir betrachten die Zufallsvariable

$$Y = (X - m_1)^2. \tag{3.3.3}$$

Sie erfüllt die Voraussetzungen des Satzes 3.3.1, da $Y \geqq 0$ ist und $E(Y)$ existiert; dabei ist

$$E(Y) = E[(X - m_1)^2] = \sigma^2.$$

Wenn wir die Konstante K gleich $k^2\sigma^2$ setzen, erhalten wir aus (3.3.2)

$$P[(X - m_1)^2 \geqq k^2\sigma^2] \leqq \frac{1}{k^2}. \tag{3.3.4}$$

Die Ungleichung (3.3.4) ist der Tschebyscheffschen Ungleichung (3.3.1) äquivalent.

Die Ungleichung (3.3.1) gilt für beliebige Zufallsvariable, die eine endliche Dispersion haben.

Wenn wir in (3.3.1) zum Beispiel den Wert $k = 3$ einsetzen, erhalten wir

$$P[|X - m_1| \geqq 3\sigma] \leqq \frac{1}{9}.$$

Die Tschebyscheffsche Ungleichung zeigt also, daß die Standardabweichung σ als Streuungsmaß dienen kann.

Aus dem nachstehenden Beispiel geht hervor, daß man in der Klasse der Zufallsvariablen, deren Moment zweiter Ordnung existiert, keine bessere Ungleichung gewinnen kann als die Tschebyscheffsche Ungleichung.

Beispiel 3.3.1. Die Zufallsvariable X möge eine Wahrscheinlichkeitsfunktion haben, die durch die Formel

$$P(X = -k) = P(X = k) = 1/2k^2,$$

$$P(X = 0) = 1 - 1/k^2$$

gegeben ist, wobei k eine gewisse positive Zahl bedeutet. Es gilt hier $m = 0$, $\sigma = 1$, und für das betrachtete k ist somit

$$P(|X - m_1| \geqq k\sigma) = P(|X| \geqq k) = P(X = -k) + P(X = k) = 1/k^2.$$

Es sei darauf hingewiesen, daß man in kleineren Klassen von Zufallsvariablen, z. B. in Klassen, deren Moment von der Ordnung $2n$ ($n > 1$) existiert, eine bessere Abschätzung erzielen kann als die Ungleichung (3.3.1) (vgl. hierzu CRAMÉR [2]).

Ein Analogon zur Tschebyscheffschen Ungleichung für zweidimensionale Zufallsvariable (vgl. Aufgabe 3.9.9) hat BERGE [1] angegeben; für Zufallsvektoren beliebig hoher endlicher Dimension stammt eine entsprechende Ungleichung von OLKIN und PRATT [1]. In 6.12.B werden wir die Ungleichung von KOLMOGOROFF angeben, die eine weitgehende Verallgemeinerung der Tschebyscheffschen Ungleichung für Summen von unabhängigen Zufallsvariablen darstellt.

3.4. Absolute Momente

Definition 3.4.1. Wir nennen den Ausdruck $E[|X|^k]$ das *absolute Moment k-ter Ordnung.*

Absolute Momente wollen wir mit β_k bezeichnen.

Wenn X eine diskrete Zufallsvariable mit den Sprungstellen x_i ist, dann ist

$$\beta_k = E[|X|^k] = \sum_i |x_i|^k p_i. \tag{3.4.1}$$

Ist X eine stetige Zufallsvariable, dann ist

$$\beta_k = E[|X|^k] = \int_{-\infty}^{\infty} |x|^k f(x)\, dx. \tag{3.4.1'}$$

Das absolute Moment gerader Ordnung ist gleich dem gewöhnlichen Moment derselben Ordnung.

Wie wir wissen (vgl. 3.2), folgt aus der Existenz des gewöhnlichen Moments k-ter Ordnung die Existenz des absoluten Moments k-ter Ordnung und auch die Existenz aller absoluten Momente einer Ordnung, die kleiner als k ist.

Wir beweisen jetzt den folgenden Satz für die absoluten Momente, der nach LJAPUNOFF [1] benannt wird:

Satz 3.4.1. *Existiert das absolute Moment n-ter Ordnung einer Zufallsvariablen X, dann gilt für $k = 1, 2, \ldots, n-1$ die Ungleichung*

$$\beta_k^{\frac{1}{k}} \leqq \beta_{k+1}^{\frac{1}{k+1}}. \tag{3.4.2}$$

Beweis. Die Zufallsvariable X sei stetig, u und v seien beliebige reelle Zahlen. Wir betrachten den nichtnegativen Ausdruck

$$\int_{-\infty}^{\infty}\left[u|x|^{\frac{k-1}{2}}+v|x|^{\frac{k+1}{2}}\right]^2 f(x)\,dx$$

$$=\int_{-\infty}^{\infty}[u^2|x|^{k-1}+2uv|x|^k+v^2|x|^{k+1}]\,f(x)\,dx=\beta_{k-1}u^2+2\beta_k uv+\beta_{k+1}v^2. \tag{3.4.3}$$

Da die quadratische Form auf der rechten Seite der Gleichung (3.4.3) ihr Vorzeichen nicht ändert, gilt

$$\beta_k^2-\beta_{k-1}\beta_{k+1}\leqq 0 \quad \text{oder} \quad \beta_k^2\leqq\beta_{k-1}\beta_{k+1}. \tag{3.4.4}$$

Erheben wir beide Seiten von (3.4.4) in die k-te Potenz, so erhalten wir

$$\beta_k^{2k}\leqq\beta_{k-1}^k\beta_{k+1}^k. \tag{3.4.5}$$

Nimmt hier k die Werte $1, 2, \ldots, n-1$ an, so folgt

$$\begin{aligned}
&\beta_1^2\leqq\beta_0\beta_2 \quad \text{mit} \quad \beta_0=\int_{-\infty}^{\infty}|x|^0 f(x)\,dx=1,\\
&\beta_2^4\leqq\beta_1^2\beta_3^2,\\
&\beta_3^6\leqq\beta_2^3\beta_4^3,\\
&\cdots\cdots\\
&\beta_{n-1}^{2(n-1)}\leqq\beta_{n-2}^{n-1}\,\beta_n^{n-1}.
\end{aligned}$$

Wenn wir die Seiten von k aufeinanderfolgenden Ungleichungen miteinander multiplizieren und beachten, daß $\beta_0=1$ ist, erhalten wir für $k=1, 2, \ldots, n-1$ die Ungleichung

$$\beta_k^{k+1}\leqq\beta_{k+1}^k. \tag{3.4.6}$$

Erheben wir noch die beiden Seiten der Ungleichung (3.4.6) in die Potenz $\dfrac{1}{k(k+1)}$, so erhalten wir (3.4.2).

Der Beweis der Ljapunoffschen Ungleichung für eine diskrete Zufallsvariable verläuft analog.

3.5. Die Lageparameter

Definition 3.5.1. Ein Wert x, der die Ungleichungen

$$P(X\leqq x)\geqq\frac{1}{2},\qquad P(X\geqq x)\geqq\frac{1}{2} \tag{3.5.1}$$

erfüllt, nennen wir *Zentralwert* oder *Mediane* und bezeichnen ihn mit $x_{\frac{1}{2}}$.

Das Ungleichungspaar (3.5.1) ist der Doppelungleichung

$$\frac{1}{2} - P(X = x) \leqq F(X) \leqq \frac{1}{2} \tag{3.5.1'}$$

äquivalent.

Ist die Zufallsvariable X stetig, so ist ein Zentralwert ein solcher Wert von X, der der Gleichung

$$F(x) = \frac{1}{2} \tag{3.5.2}$$

genügt. Es kann vorkommen, daß mehrere x-Werte die Beziehung (3.5.1) bzw. (3.5.2) erfüllen. In diesem Fall nennen wir jeden von ihnen Zentralwert.

Beispiel 3.5.1. Die Zufallsvariable X möge die Werte 0 und 1 mit den Wahrscheinlichkeiten

$$P(X = 0) = \frac{1}{5}, \qquad P(X = 1) = \frac{4}{5}$$

annehmen. Der Zentralwert ist der Punkt $x = 1$; denn es ist

$$P(X \leqq 1) = P(X = 0) + P(X = 1) = 1 > \frac{1}{2},$$

$$P(X \geqq 1) = P(X = 1) = \frac{4}{5} > \frac{1}{2}.$$

Beispiel 3.5.2. Die Zufallsvariable X sei stetig mit der durch

$$f(x) = \begin{cases} 0 & \text{für} \quad x < 0, \\ \cos x & \text{für} \quad 0 \leqq x \leqq \frac{\pi}{2}, \\ 0 & \text{für} \quad x > \frac{\pi}{2} \end{cases}$$

gegebenen Dichte. Den Zentralwert von X erhalten wir aus den Gleichungen (3.5.2):

$$\int_{-\infty}^{x_{1/2}} f(x)\,dx = \int_{0}^{x_{1/2}} \cos x\,dx = \sin x \Big|_0^{x_{1/2}} = \sin x_{1/2} = \frac{1}{2},$$

und zwar ist

$$x_{1/2} = \frac{\pi}{6}.$$

Beispiel 3.5.3. Die Zufallsvariable X möge die drei Werte $x_1 = -1$, $x_2 = 0$ und $x_3 = 1$ mit den Wahrscheinlichkeiten $P(X = -1) = P(X = 0) = \frac{1}{4}$, $P(X = 1) = \frac{1}{2}$ annehmen.

Abb. 3.5.1 stellt die Verteilungsfunktion dieser Zufallsvariablen dar.

Jeder Wert x aus dem Intervall $[0, 1]$ ist hier Zentralwert, da für jedes derartige x die Beziehungen

$$P(X \leqq x) \geqq \frac{1}{2}, \qquad P(X \geqq x) \geqq \frac{1}{2}$$

gelten.

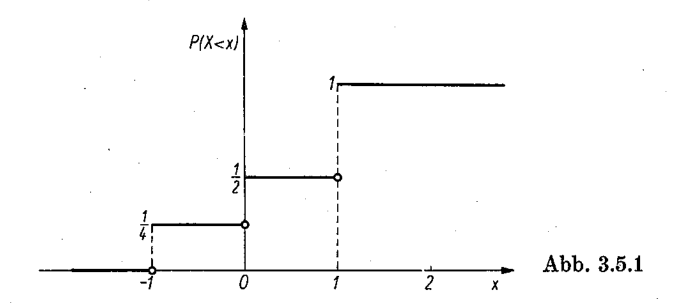


Abb. 3.5.1

Der Zentralwert gehört zu einer Gruppe von Parametern, die man als *Quantile* bezeichnet.

Definition 3.5.2. Einen Wert x, der den Ungleichungen

$$P(X \leqq x) \geqq p, \quad P(X \geqq x) \geqq 1 - p \qquad (0 < p < 1) \tag{3.5.3}$$

genügt, nennen wir *Quantil p-ter Ordnung* und bezeichnen ihn mit x_p.

Das Ungleichungspaar (3.5.3) ist der Doppelungleichung

$$p - P(X = x) \leqq F(x) \leqq p \tag{3.5.3'}$$

äquivalent.

Hat die Zufallsvariable X eine stetige Verteilung, so ist das Quantil p-ter Ordnung derjenige Wert von x, der die Gleichung

$$F(x) = p \tag{3.5.4}$$

erfüllt. Es kann vorkommen, daß den Ungleichungen (3.5.3) bzw. der Gleichung (3.5.4) mehrere Werte x genügen. In diesem Fall bezeichnen wir jeden von ihnen als Quantil p-ter Ordnung.

Die Quantile und ihre Funktionen werden *Lageparameter* genannt.

Beispiel 3.5.4. Die Zufallsvariable X sei normal verteilt mit der durch

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

gegebenen Dichte. Gesucht ist der Punkt x, für den die Gleichung $F(x) = \frac{1}{10}$ gilt.

Wir benutzen die Tafel III der Normalverteilung am Schluß des Buches und finden

$$x_{\frac{1}{10}} \approx -1{,}28.$$

Einige einfache Funktionen der Quantile können gleichfalls als Streuungsmaße dienen. Ein solches Streuungsmaß stellt zum Beispiel der durch

$$\frac{x_{\frac{3}{4}} - x_{\frac{1}{4}}}{2} \tag{3.5.5}$$

definierte *Viertelwertsabstand* dar.

Wenn die Menge aller möglichen Werte einer Zufallsvariablen nach beiden Seiten beschränkt ist, existiert für diese Menge eine endliche obere und eine endliche untere Grenze. Im Fall einer beschränkten diskreten Zufallsvariablen mit endlich vielen Sprungstellen stimmt die obere bzw. die untere Grenze mit dem größten bzw. dem kleinsten Wert überein, den die Zufallsvariable annehmen kann.

Ein wichtiger Lageparameter, der ebenfalls als Streuungsmaß dient, ist die *Variationsbreite.*

Sind a bzw. b die untere bzw. obere Grenze, so wird die Variationsbreite durch

$$d = b - a \tag{3.5.6}$$

definiert.

In Beispiel 3.5.2 ist die Variationsbreite gleich $\frac{\pi}{2}$, während sie im Beispiel 3.5.3 gleich 2 ist.

3.6. Die Momente einer mehrdimensionalen Zufallsvariablen

A. Es sei (X, Y) eine zweidimensionale Zufallsvariable. Wir betrachten die eindeutige Funktion $g(X, Y)$ der Zufallsvariablen (X, Y).

Definition 3.6.1. Es sei (X, Y) eine diskrete Zufallsvariable mit den Sprungstellen (x_i, y_k) und den Sprunghöhen p_{ik}. Die Funktion $g(x, y)$ sei im Riemannschen Sinne integrierbar. Ist die Ungleichung

$$\sum_{i,k} p_{ik} \, |g(x_i, y_k)| < \infty \tag{3.6.1'}$$

erfüllt, so nennen wir die Reihe

$$\sum_{i,k} p_{ik} \, g(x_i, y_k) \tag{3.6.1}$$

den *Mittelwert* $E[g(X, Y)]$ *der Zufallsvariablen* $g(X, Y)$.

Definition 3.6.2. Es sei (X, Y) eine stetige Zufallsvariable mit der Dichtefunktion $f(x, y)$. Ist die Ungleichung

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| \, f(x, y) \, dx \, dy < \infty \tag{3.6.2'}$$

erfüllt, so nennen wir das Integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \, f(x, y) \, dx \, dy \tag{3.6.2}$$

den *Mittelwert* $E[g(X, Y)]$ *der Zufallsvariablen* $g(X, Y)$.

Definition 3.6.3. Den Mittelwert der Funktion $g(X, Y) = X^l Y^n$ nennen wir *Moment der Zufallsvariablen* (X, Y) und schreiben

$$m_{ln} = E(X^l Y^n);$$

dabei sind l und n nichtnegative ganze Zahlen. Die Summe $l + n$ nennt man die *Ordnung des Moments* m_{ln}.

Ist also (X, Y) eine diskrete Zufallsvariable mit den Sprungstellen (x_i, y_k) und den Sprunghöhen p_{ik}, so ist nach (3.6.1)

$$m_{ln} = \sum_{i,k} p_{ik} x_i^l y_k^n. \tag{3.6.3}$$

Ist dagegen (X, Y) eine stetige Zufallsvariable mit der Dichte $f(x, y)$, so erhalten wir nach (3.6.2)

$$m_{ln} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^l y^n f(x, y) \, dx \, dy. \tag{3.6.4}$$

Wir erinnern daran, daß für die Existenz des Moments m_{ln} die absolute Konvergenz der Reihe (3.6.3) bzw. des Integrals (3.6.4) notwendig ist.

Es können zwei Momente erster Ordnung existieren: m_{10} und m_{01}. Wie man leicht einsieht, sind dies die Mittelwerte der Randverteilungen der Zufallsvariablen X und Y, also

$$m_{10} = E(X), \quad m_{01} = E(Y). \tag{3.6.5}$$

Es können drei Momente zweiter Ordnung existieren: m_{20}, m_{02}, m_{11}, und es bestehen die Beziehungen

$$m_{20} = E(X^2), \quad m_{02} = E(Y^2). \tag{3.6.6}$$

Die zentralen Momente definiert man ähnlich wie bei einer eindimensionalen Verteilung. Wir wollen sie mit μ_{ln} bezeichnen und erhalten

$$\mu_{ln} = E[(X - m_{10})^l (Y - m_{01})^n]. \tag{3.6.7}$$

Insbesondere erhalten wir

$$\mu_{10} = E(X - m_{10}) = 0, \tag{3.6.8}$$
$$\mu_{01} = E(Y - m_{01}) = 0,$$
$$\mu_{20} = E[(X - m_{10})^2] = \sigma_1^2, \tag{3.6.9}$$
$$\mu_{02} = E[(Y - m_{01})^2] = \sigma_2^2.$$

Dabei sind σ_1 bzw. σ_2 die Standardabweichungen der Zufallsvariablen X bzw. Y. Das zentrale Moment μ_{11} nennen wir die *Kovarianz* und schreiben dafür cov (X, Y).

Zwischen den gewöhnlichen und den zentralen Momenten bestehen ähnliche Beziehungen wie bei einer eindimensionalen Verteilung:

$$\begin{aligned} \mu_{20} &= m_{20} - m_{10}^2, \\ \mu_{02} &= m_{02} - m_{01}^2. \end{aligned} \tag{3.6.10}$$

Für die Kovarianz erhalten wir

$$\begin{aligned} \mu_{11} &= E[(X - m_{10})(Y - m_{01})] \\ &= E(XY) - m_{10}E(Y) - m_{01}E(X) + m_{10}m_{01} \\ &= m_{11} - m_{10}m_{01} - m_{01}m_{10} + m_{10}m_{01} = m_{11} - m_{10}m_{01}. \end{aligned} \tag{3.6.11}$$

B. Wir wollen uns jetzt mit dem Mittelwert der Summe und des Produkts von Zufallsvariablen beschäftigen. X und Y seien zwei (abhängige oder unabhängige) Zufallsvariable. Wir setzen die Existenz der Mittelwerte $E(X)$ und $E(Y)$ voraus, betrachten die Zufallsvariable $Z = X + Y$ und wollen den Mittelwert $E(Z)$ berechnen.

Es sei (X, Y) eine diskrete Zufallsvariable mit den Sprungstellen (x_i, y_k) und den Sprunghöhen p_{ik}. Wenn wir in (3.6.1) $Z = X + Y = g(X, Y)$ einsetzen und berücksichtigen, daß $E(X)$ und $E(Y)$ existieren, erhalten wir

$$\begin{aligned} E(Z) &= \sum_{i,k} p_{ik}(x_i + y_k) = \sum_i x_i \sum_k p_{ik} + \sum_k y_k \sum_i p_{ik} \\ &= \sum_i x_i p_{i\cdot} + \sum_k y_k p_{\cdot k} = E(X) + E(Y). \end{aligned} \tag{3.6.12}$$

Es sei jetzt (X, Y) eine stetige Zufallsvariable mit der Dichtefunktion $f(x, y)$. In diesem Fall erhalten wir aus (3.6.2)

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x, y)\, dy\, dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} x f_1(x)\, dx + \int_{-\infty}^{\infty} y f_2(y)\, dy = E(X) + E(Y). \end{aligned} \tag{3.6.13}$$

Mit Hilfe vollständiger Induktion kann man die Formeln (3.6.12) und (3.6.13) auf endlich viele Zufallsvariable verallgemeinern. Es gilt also der

Satz 3.6.1. *Der Mittelwert einer Summe endlich vieler Zufallsvariabler ist gleich der Summe der Mittelwerte dieser Zufallsvariablen.*

Für das Produkt von Zufallsvariablen gilt der folgende

Satz 3.6.2. *Der Mittelwert des Produkts von endlich vielen unabhängigen Zufallsvariablen ist gleich dem Produkt der Mittelwerte dieser Zufallsvariablen.*

Beweis. Zum Beweis beschränken wir uns auf ein Produkt von zwei Zufallsvariablen, da man den Satz durch vollständige Induktion auf eine beliebige Faktorenanzahl ausdehnen kann.

Es sei (X, Y) eine diskrete Zufallsvariable mit den Mittelwerten $E(X)$ und $E(Y)$, die existieren mögen. Aus (3.6.1) erhalten wir wegen der Unabhängigkeit von X und Y

$$\begin{aligned} E(XY) &= \sum_{i,k} p_{ik} x_i y_k = \sum_{i,k} p_{i\cdot} p_{\cdot k} x_i y_k \\ &= \sum_i p_{i\cdot} x_i \sum_k p_{\cdot k} y_k = E(X)\,E(Y). \end{aligned} \tag{3.6.14}$$

Analog erhalten wir, wenn (X, Y) eine stetige Zufallsvariable ist,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y)\,dx\,dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_1(x) f_2(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} x f_1(x)\,dx \int_{-\infty}^{\infty} y f_2(y)\,dy = E(X)\,E(Y). \end{aligned} \tag{3.6.15}$$

Damit ist Satz 3.6.2 bewiesen (wegen des Beweises des umgekehrten Satzes vgl. den Anhang).

Aus der Formel (3.6.11) für die Kovarianz und aus Satz 3.6.2 folgt, daß die Kovarianz zweier Zufallsvariablen gleich Null ist. Es ist nämlich

$$\mu_{11} = E(XY) - E(X)\,E(Y) = 0.$$

Daraus, daß die Kovarianz zweier Zufallsvariablen gleich Null ist, folgt aber noch nicht die Unabhängigkeit der Zufallsvariablen.

Es seien X und Y zwei Zufallsvariable mit endlichen Dispersionen $D^2(X)$ und $D^2(Y)$. Wir betrachten die Zufallsvariable

$$Z = X + Y.$$

Die Dispersion von Z ist gleich

$$\begin{aligned} D^2(Z) = D^2(X + Y) &= E[(X + Y)^2] - [E(X + Y)]^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - [E(Y)]^2 - 2E(X)\,E(Y) \\ &= E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2 + 2E(XY) - 2E(X)\,E(Y). \end{aligned}$$

Schließlich erhalten wir

$$D^2(X + Y) = D^2(X) + D^2(Y) + 2E(XY) - 2E(X)E(Y). \qquad (3.6.16)$$

Wenn X und Y unabhängig sind, erhalten wir auf Grund von (3.6.14)

$$D^2(Z) = D^2(X) + D^2(Y). \qquad (3.6.17)$$

Dieses Ergebnis kann auf die Summe endlich vieler Zufallsvariabler ausgedehnt werden:

Satz 3.6.3. *Die Dispersion der Summe von endlich vielen unabhängigen Zufallsvariablen ist gleich der Summe der Dispersionen dieser Zufallsvariablen.*

C. Wir leiten nun die Formeln für die Momente der bedingten Verteilung einer der Zufallsvariablen X, Y unter der Bedingung her, daß die andere einen festen Wert annimmt.

Die Zufallsvariable (X, Y) sei diskret mit den Sprungstellen (x_i, y_k) und den Sprunghöhen p_{ik}. Wir bezeichnen mit $E(Y^l|X = x_i)$ den bedingten Mittelwert der Zufallsvariablen Y^l $(l = 1, 2, \ldots)$, wenn $X = x_i$ ist, und erhalten aus den Formeln (2.7.2) und (3.2.1)

$$E(Y^l|X = x_i) = \sum_k y_k^l \frac{p_{ik}}{p_{i\cdot}}. \qquad (3.6.18)$$

Analog erhalten wir für den bedingten Mittelwert der Zufallsvariablen X^l $(l = 1, 2, \ldots)$, wenn $Y = y_k$ ist, die Formel

$$E(X^l|Y = y_k) = \sum_i x_i^l \frac{p_{ik}}{p_{\cdot k}}. \qquad (3.6.19)$$

Es sei (X, Y) eine stetige Zufallsvariable mit der Dichtefunktion $f(x, y)$. Aus den Formeln (2.7.6) und (3.2.2) ergibt sich

$$E(Y^l|X = x) = \int_{-\infty}^{\infty} y^l \frac{f(x, y)}{f_1(x)} \, dy \qquad (3.6.20)$$

und ebenso

$$E(X^l|Y = y) = \int_{-\infty}^{\infty} x^l \frac{f(x, y)}{f_2(y)} \, dx. \qquad (3.6.21)$$

Die Formeln (3.6.18) bis (3.6.21) sind selbstverständlich nur dann sinnvoll, wenn die betrachteten Momente existieren.

Die bedingten Momente der einen Zufallsvariablen sind konstante Größen, sofern die Werte der anderen Zufallsvariablen festgelegt sind. So ist zum Beispiel für ein festes x_i das Moment $E(Y^l|X = x_i)$ konstant. Aber x_i ist ein Wert der Zufallsvariablen X; wir können daher $E(Y^l|x)$ als Zufallsvariable auffassen, die

die Werte $E(Y^l|X = x_i)$ mit der Wahrscheinlichkeit p_i $(i = 1, 2, \ldots)$ annehmen. Berücksichtigen wir die Formel (3.6.18), so finden wir für eine beliebige Teilmenge S der Sprungstellen x_i der Zufallsvariablen X die Gleichung (vgl. 2.7.C)

$$\begin{aligned} E(Y^l|X \in S) &= \sum_k y_k^l P(Y = y_k|X \in S) \\ &= \sum_k y_k^l \Big(\sum_{x_i \in S} p_{ik} \Big/ \sum_{x_i \in S} p_{i\cdot} \Big) \\ &= \sum_{x_i \in S} p_{i\cdot} \sum_k \frac{y_k^l p_{ik}}{p_{i\cdot}} \Big/ \sum_{x_i \in S} p_{i\cdot} \\ &= \sum_{x_i \in S} \Big(p_{i\cdot} \Big/ \sum_{x_i \in S} p_{i\cdot} \Big) E(Y^l|X = x_i) \\ &= E[E(Y^l|X = x_i)|X \in S]. \end{aligned} \tag{3.6.22}$$

Nehmen wir insbesondere für S die Menge aller Sprungstellen x_i der Zufallsvariablen X, so erhalten wir

$$E(Y^l) = \sum_i \sum_k y_k^l p_{ik} = \sum_i p_{i\cdot} E(Y^l|X = x_i) = E[E(Y^l|x)]. \tag{3.6.22'}$$

Ähnlich erhalten wir im Fall einer stetigen Zufallsvariablen aus der Formel (3.6.20) für eine beliebige Borelsche Menge S auf der x-Achse, für die $P(X \in S) > 0$ gilt, die Gleichung

$$\begin{aligned} E(Y^l|X \in S) &= \int_{-\infty}^{\infty} y^l \int_S f(x, y)\, dx\, dy / P(X \in S) \\ &= \int_S f_1(x) \int_{-\infty}^{\infty} y^l \frac{f(x, y)}{f_1(x)}\, dy\, dx / P(X \in S) \\ &= \int_S \frac{f_1(x)}{P(X \in S)} E(Y^l|X = x)\, dx \\ &= E[E(Y^l|X = x)|X \in S]. \end{aligned} \tag{3.6.23}$$

Setzen wir $S = (-\infty, \infty)$, so ergibt sich

$$\begin{aligned} E(Y^l) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^l f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} f_1(x)\, E(Y^l|X = x)\, dx \\ &= E[E(Y^l|x)]. \end{aligned} \tag{3.6.23'}$$

Die Formeln (3.6.22) und (3.6.23) lassen sich auch durch die eine Formel

$$E(Y^l \mid X \in S) = E[E(Y^l \mid X = x) \mid X \in S] \tag{3.6.24}$$

darstellen, während wir (3.6.22') und (3.6.23') durch

$$E(Y^l) = E[E(Y^l \mid x)] \tag{3.6.24'}$$

darstellen können.

Die gewonnenen Formeln eröffnen uns einen neuen Zugang zum Problem der bedingten Wahrscheinlichkeit. Es sei E die Menge der Elementarereignisse e und Z ein Borelscher Körper zufälliger Ereignisse, die Untermengen von E sind. Ferner sei $A \in Z$ ein festes zufälliges Ereignis. Nun definieren wir eine Zufallsvariable Y folgendermaßen:

$$Y(e) = \begin{cases} 1 & \text{für} \quad e \in A, \\ 0 & \text{für} \quad e \notin A; \end{cases}$$

es ist somit $P(Y = 1) = P(A)$. Sodann sei X eine beliebige andere, in derselben Menge E der Elementarereignisse definierte Zufallsvariable mit der Eigenschaft, daß die rechte Seite von (3.6.24) existiert. Die Beziehungen (3.6.24) bzw. (3.6.24') erhalten nunmehr die Form

$$P(A \mid X \in S) = E[P(A \mid X = x) \mid X \in S] \tag{3.6.24''}$$

bzw.

$$P(A) = E[P(A \mid x)], \tag{3.6.24'''}$$

und es ist daher $P(A \mid x)$ eine Zufallsvariable.

Man beachte, daß (3.6.24), (3.6.24'), (3.6.24'') und (3.6.24''') unter der Voraussetzung hergeleitet wurden, daß die Zufallsvariable (X, Y) entweder diskret oder stetig ist und daß die in 2.7 formulierten zusätzlichen Bedingungen erfüllt sind. Im allgemeinen Fall jedoch dienen die Gleichungen (3.6.24) bzw. (3.6.24'') (vgl. Anhang) als Definitionen des bedingten Mittelwertes bzw. der bedingten Wahrscheinlichkeit. Diese Definitionen verdanken wir KOLMOGOROFF [7], der zugleich gezeigt hat, daß, falls $E(Y^l)$ existiert, die Gleichung (3.6.24) die Zufallsvariable $E(Y^l \mid x)$ bis auf Äquivalenz eindeutig bestimmt, d. h., genügt irgendeine andere Zufallsvariable U, anstelle von $E(Y^l \mid x)$ in (3.6.24) eingesetzt, dieser Gleichung, so gilt

$$P[U - E(Y^l \mid x) = 0] = 1.$$

Insbesondere bestimmt somit die Gleichung (3.6.24'') bis auf Äquivalenz eindeutig die bedingte Wahrscheinlichkeit $P(A \mid x)$.

Eine weitere Entwicklung dieser Gedankengänge findet der Leser in der Monographie von DOOB [5].

D. Bis jetzt haben wir uns in 3.6 auf zweidimensionale Zufallsvariable beschränkt; die Ausdehnung der eingeführten Begriffe auf mehr als zwei Dimensionen bereitet jedoch keine Schwierigkeiten. Wir wollen hier nur noch einige Bemerkungen über den bedingten Mittelwert in m-dimensionalen ($m > 2$) Verteilungen machen.

Wir betrachten die stetige Zufallsvariable $(X_1, X_2, \ldots, X_m)$ und nehmen an, daß die Dichtefunktion $f(x_1, x_2, \ldots, x_m)$ überall stetig und die Dichtefunktion der Randverteilung

$$\int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_m)\, dx_1 = g(x_2, \ldots, x_m)$$

der Zufallsvariablen $(X_2, \ldots, X_m)$ überall stetig und positiv ist. Wir können dann die bedingten Momente l-ter Ordnung ($l = 1, 2, 3, \ldots$)

$$E(X_1^l \mid X_2 = x_2, \ldots, X_m = x_m) = \int_{-\infty}^{\infty} x_1^l \frac{f(x_1, x_2, \ldots, x_m)}{g(x_2, \ldots, x_m)}\, dx_1 \qquad (3.6.25)$$

betrachten.

Ebenso kann man auch die bedingten Momente der übrigen Zufallsvariablen und die mehrdimensionalen bedingten Momente irgendeiner Gruppe von Zufallsvariablen bilden.

Für diskrete Zufallsvariable ist der Gedankengang ganz analog.

E. Falls in einer zweidimensionalen Verteilung die Standardabweichungen σ_1 und σ_2 existieren und von Null verschieden sind, kann man einen sehr wichtigen Parameter definieren, der die zweidimensionale Verteilung charakterisiert, nämlich den *Korrelationskoeffizienten* ϱ:

$$\varrho = \frac{E[(X - m_{10})(Y - m_{01})]}{\sqrt{E[(X - m_{10})^2]}\,\sqrt{E[(Y - m_{01})^2]}} = \frac{\mu_{11}}{\sigma_1 \sigma_2}. \qquad (3.6.26)$$

Eine grundlegende Eigenschaft des Korrelationskoeffizienten ϱ gibt der folgende Satz an.

Satz 3.6.4. *Der durch* (3.6.26) *definierte Korrelationskoeffizient* ϱ *erfüllt die Ungleichung*

$$-1 \leqq \varrho \leqq 1. \qquad (3.6.27)$$

Beweis. Wir betrachten für beliebige reelle t und u den nichtnegativen Ausdruck

$$\begin{aligned} &E\{[t(X - m_{10}) + u(Y - m_{01})]^2\} \\ &= t^2 E[(X - m_{10})^2] + 2tuE[(X - m_{10})(Y - m_{01})] + u^2 E[(Y - m_{01})^2] \\ &= t^2\sigma_1^2 + 2tu\mu_{11} + u^2\sigma_2^2. \end{aligned} \qquad (3.6.28)$$

Da die linke Seite des Ausdrucks (3.6.28) niemals negativ ist, müssen die Ungleichungen

$$\mu_{11}^2 - \sigma_1^2\sigma_2^2 \leqq 0 \quad \text{oder} \quad -\sigma_1\sigma_2 \leqq \mu_{11} \leqq \sigma_1\sigma_2$$

gelten, also

$$-1 \leqq \frac{\mu_{11}}{\sigma_1\sigma_2} \leqq 1 .$$

Nach der Definition des Korrelationskoeffizienten folgt daraus unmittelbar die Ungleichung (3.6.27).

Sind die Zufallsvariablen X und Y unabhängig, so ist $\mu_{11} = 0$, also auch $\varrho = 0$. Wie oben schon erwähnt wurde, ist die Umkehrung dieses Satzes nicht richtig. Wenn $\varrho = 0$ ist, so sagen wir, daß die Zufallsvariablen *unkorreliert* sind (zum Unterschied von „unabhängig").

Wie man bemerkt, ist Satz 3.6.3 auch dann richtig, wenn je zwei der dort vorkommenden Zufallsvariablen unkorreliert sind.

Eine Antwort auf die Frage, was es bedeutet, wenn der Korrelationskoeffizient ϱ^2 gleich 1 ist, gibt der folgende Satz:

Satz 3.6.5. *Die Beziehung*

$$P(Y = aX + b) = 1 \tag{3.6.29}$$

gilt genau dann, wenn $\varrho^2 = 1$ *ist.*

Ehe wir diesen Satz beweisen, wollen wir ihn interpretieren. Nach diesem Satz ist die Wahrscheinlichkeit dafür, daß die Zufallsvariable (X, Y) Werte annimmt, die auf einer Geraden der x, y-Ebene liegen, gleich Eins, wenn $\varrho^2 = 1$ ist. Würden wir diese Gerade als Koordinatenachse wählen, so hätten wir es mit einer eindimensionalen Zufallsvariablen zu tun. Man veranschaulicht diesen Tatbestand, indem man sagt, die ganze „Wahrscheinlichkeitsmasse" liege auf einer Geraden.

Beweis. Wir nehmen an, die Beziehung (3.6.29) sei erfüllt, es sei also

$$P(Y \neq aX + b) = 0 .$$

Nach (3.6.24) erhalten wir

$$\begin{aligned} m_{01} &= E(Y) \\ &= P(Y = aX + b)\, E(Y \mid Y = aX + b) \\ &\quad + P(Y \neq aX + b)\, E(Y \mid Y \neq aX + b) \\ &= E(aX + b) = a m_{10} + b , \end{aligned}$$

$$\begin{aligned}\sigma_2^2 &= E[(Y - m_{01})^2] \\ &= P(Y = aX + b)\, E[(Y - m_{01})^2 \mid Y = aX + b] \\ &\quad + P(Y \neq aX + b)\, E[(Y - m_{01})^2 \mid Y \neq aX + b] \\ &= E[(aX + b - m_{01})^2] = E[a^2(X - m_{10})^2] \\ &= a^2 E[(X - m_{10})^2] = a^2\sigma_1^2 .\end{aligned}$$

Analog erhalten wir $\mu_{11} = a\sigma_1^2$. Aus diesen Gleichungen folgt $\varrho^2 = 1$.

Es sei nun $\varrho^2 = 1$. Aus der Definition des Korrelationskoeffizienten erhalten wir

$$\mu_{20}\mu_{02} - \mu_{11}^2 = 0 .$$

Die quadratische Form (3.6.28) nimmt also den Wert Null für ein gewisses Zahlenpaar $t = t_0$, $u = u_0$ an, wobei t_0 und u_0 nicht gleichzeitig Null sind. Für diese Werte t_0, u_0 ist also

$$E\{[t_0(X - m_{10}) + u_0(Y - m_{01})]^2\} = 0 .$$

Diese Gleichung ist nur dann erfüllt, wenn

$$P[t_0(X - m_{10}) + u_0(Y - m_{01}) = 0] = 1$$

gilt. Wenn wir $u_0 \neq 0$ annehmen, dann folgt daraus

$$P\left(Y = -\frac{t_0}{u_0}X + \frac{m_{01}u_0 + m_{10}t_0}{u_0}\right) = 1 .$$

Der Satz 3.6.5 ist also vollständig bewiesen.

Aus diesem Satz folgt, daß der Fall $\varrho^2 = 1$ den größten Gegensatz zur Unabhängigkeit darstellt, also der Fall extremer Abhängigkeit ist. Ist $\varrho = 1$, dann ist die Änderungsrichtung der beiden Veränderlichen dieselbe, d. h., größeren Werten der einen Veränderlichen entsprechen auch größere Werte der anderen und umgekehrt. Ist dagegen $\varrho = -1$, so ist die Änderungsrichtung entgegengesetzt.

Beispiel 3.6.1. Die Zufallsvariablen X und Y haben die gemeinsame Dichte

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2 - 2xy + 2y^2}{2}} .$$

Dieser Ausdruck ist eine Wahrscheinlichkeitsdichte; die Funktion $f(x, y)$ ist nämlich positiv, und außerdem ist

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2 - 2xy + 2y^2}{2}}\,dx\,dy$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{(x-y)^2}{2}}\,dx\,dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy = 1 .$$

Die Verteilung der betrachteten Zufallsvariablen ist ein Spezialfall der zweidimensionalen Normalverteilung (siehe 5.11).

Wir berechnen die Momente erster und zweiter Ordnung:

$$m_{10} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y)\, dx\, dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-y)^2}{2}}\, dx\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2}}\, dy = 0,$$

$$m_{01} = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{4}} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} y e^{-\left(y-\frac{x}{2}\right)^2}\, dy\, dx = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{x}{2} e^{-\frac{x^2}{4}}\, dx = 0.$$

Also erhalten wir

$$\mu_{11} = m_{11} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-y)^2}{2}}\, dx\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}}\, dy = 1,$$

$$\mu_{20} = m_{20} = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{4}} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\left(y-\frac{x}{2}\right)^2}\, dy\, dx$$

$$= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{4}}\, dx = 2,$$

$$\mu_{02} = m_{02} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{2}}\, dx\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^3}{2}}\, dy = 1.$$

Daraus bekommen wir als Korrelationskoeffizienten

$$\varrho = \frac{\mu_{11}}{\sigma_1 \sigma_2} = \frac{1}{\sqrt{2}}.$$

Neben dem Korrelationskoeffizienten gibt es viele andere Maße für die Abhängigkeit zwischen zwei Zufallsvariablen. Ausführliche Informationen hierüber findet man in den Büchern von EZEKIEL [1] und M. G. KENDALL [1]. Unter den neueren Arbeiten auf diesem Gebiet nennen wir die von GEBELEIN [1], HIRSCHFELD [1] sowie RÉNYI [8], [9]. In den letzten beiden Arbeiten werden einige wichtige Abhängigkeitsmaße zwischen zwei Zufallsvariablen miteinander verglichen.

F. Wir wollen uns jetzt den n-dimensionalen Zufallsvariablen $(X_1, X_2, \ldots, X_n)$ zuwenden. Wir setzen voraus, daß die Dispersionen σ_i^2 $(i = 1, 2, \ldots, n)$ der Zufallsvariablen X_i existieren und positiv sind und daß die Kovarianzen aller Paare von Zufallsvariablen existieren. Mit λ_{ik} bzw. ϱ_{ik} bezeichnen wir die Kovarianz bzw. den Korrelationskoeffizienten der Zufallsvariablen X_i und X_k, also ist $\lambda_{ii} = \sigma_i^2$.

Die symmetrische Matrix

$$M = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \lambda_{n1} & \lambda_{n2} & \ldots & \lambda_{nn} \end{pmatrix}, \tag{3.6.30}$$

deren Elemente die Momente λ_{ik} zweiter Ordnung sind, nennen wir *Momentenmatrix.* Mit $|M|$ bezeichnen wir die Determinante der Matrix M.

Eine Verallgemeinerung des Satzes 3.6.5 auf n Zufallsvariable wäre ein Satz, der zu entscheiden erlaubt, wann zwischen den Zufallsvariablen $X_1, X_2, \ldots, X_n$ irgendwelche linearen Beziehungen bestehen. Das gewünschte Kriterium gibt der folgende Satz.

Satz 3.6.6. *Haben die Zufallsvariablen $X_1, \ldots, X_n$ endliche Dispersionen und Kovarianzen, so ist die Bedingung*

$$|M| = 0$$

notwendig und hinreichend dafür, daß mit der Wahrscheinlichkeit Eins unter den Zufallsvariablen mindestens eine lineare Beziehung besteht.

Der Beweis von Satz 3.6.6 verläuft ebenso wie der des Satzes 3.6.5, nur mit dem Unterschied, daß hier die nichtnegativ definite quadratische Form

$$E\left\{\left[\sum_{i=1}^{n} t_i\big(X_i - E(X_i)\big)\right]^2\right\} = \sum_{i,k=1}^{n} \lambda_{ik} t_i t_k \geqq 0 \tag{3.6.31}$$

an die Stelle des Ausdrucks (3.6.28) tritt.

Die Interpretation ist ähnlich wie bei Satz 3.6.5. Ist nämlich $|M| = 0$, dann liegt die ganze „Wahrscheinlichkeitsmasse“ in einer Hyperebene von kleinerer Dimension als n.

G. Im Zusammenhang mit dem Satz 3.6.6 geben wir die folgende Definition:

Definition 3.6.4. Genügen die Komponenten $X_1, \ldots, X_n$ eines Zufallsvektors $(X_1, \ldots, X_n)$ wenigstens einer linearen Beziehung mit der Wahrscheinlichkeit 1, so nennt man die Verteilung von $(X_1, \ldots, X_n)$ *ausgeartet.*

Ist die Determinante $|M| \neq 0$, so ist die Verteilung von $(X_1, \ldots, X_n)$ *nicht ausgeartet.*

Definition 3.6.5. Die Determinante $|M|$ nennen wir die *verallgemeinerte Varianz.*

Um diese Bezeichnung zu begründen, zeigen wir, daß $|M|$ als Streuungsmaß dienen kann. Zunächst stellen wir fest, daß die quadratische Form (3.6.31) wegen $|M| \neq 0$ nicht negativ ist. Ferner sehen wir, daß wegen $\lambda_{ik} = \lambda_{ki}$ die Matrix M symmetrisch und positiv definit ist. Hieraus geht hervor, daß $|M| \leqq \sigma_1^2 \cdots \sigma_n^2$ ist. Die Determinante $|M|$ hat somit ihren größten Wert für $\lambda_{ik} = 0$ $(i \neq k)$, d. h. für den Fall, daß die unabhängigen Variablen $X_1, \ldots, X_n$ nicht korreliert sind. Andererseits hat das Produkt $\sigma_1^2 \cdots \sigma_n^2$ einen umso kleineren Wert, je kleiner die σ_i^2 und damit je weniger gestreut die Randverteilungen der Variablen X_i sind. Das genannte Produkt ist aber die obere Grenze der verallgemeinerten Varianz. In dem äußersten Fall, daß wenigstens eines der σ_i^2 gegen Null strebt, artet wegen Satz 3.6.6 die Verteilung von $(X_1, \ldots, X_n)$ aus, d. h., sie wird zur Verteilung einer Zufallsvariablen von weniger als n Dimensionen. Es ist durchaus natürlich anzunehmen, daß eine solche Verteilung in einem n-dimensionalen Raum äußerst konzentriert erscheint.

Der Begriff der verallgemeinerten Varianz wurde von Wilks [1] eingeführt.

Wir betrachten nun die Matrix R der Korrelationskoeffizienten ϱ_{ik} und nehmen dabei $\varrho_{ii} = 1$ an:

$$R = \begin{pmatrix} 1 & \varrho_{12} & \cdots & \varrho_{1n} \\ \varrho_{21} & 1 & \cdots & \varrho_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \varrho_{n1} & \varrho_{n2} & \cdots & 1 \end{pmatrix}.$$

Wegen $\varrho_{ik} = \dfrac{\lambda_{ik}}{\sigma_i \sigma_k}$ erhalten wir für die Determinanten $|R|$ und $|M|$ die Beziehung

$$|M| = \sigma_1^2 \cdots \sigma_n^2 |R|. \tag{3.6.34}$$

Die Matrix R ist ebenfalls symmetrisch und positiv definit, und ihre Determinante genügt der Ungleichung $|R| \leqq 1$. Ihren größten Wert 1 nimmt die Determinante $|R|$ an, wenn sämtliche Korrelationskoeffizienten $\varrho_{ik} = 0$ sind. Ihren Minimalwert 0 nimmt sie an, wenn wenigstens einer ihrer Korrelationskoeffizienten den Wert 1 annimmt und die Verteilung somit ausgeartet ist. Hieraus geht hervor, daß der Wert der Determinante $|R|$ als Maß der Aus-

artung der Verteilung des Vektors $(X_1, \ldots, X_n)$ dienen kann. Ein solches Maß hat FRISCH [1] durch die folgende Definition eingeführt:

Definition 3.6.6. Der Ausdruck $\sqrt{|R|}$ heißt *Dispersionskoeffizient.*

3.7. Die Regression erster Art

A. Es sei (X, Y) eine zweidimensionale diskrete Zufallsvariable mit den Sprungstellen (x_i, y_k) und den Sprunghöhen p_{ik}, während durch $p_{i\cdot}$ bzw. $p_{\cdot k}$ wie bisher die Randverteilungswahrscheinlichkeiten der Zufallsvariablen X bzw. Y bezeichnet werden.

Wir betrachten die bedingten Mittelwerte von X und Y, die wir mit $m_1(y_k)$ und $m_2(x_i)$ bezeichnen wollen und die wir aus den Formeln (3.6.18) und (3.6.19) erhalten, wenn wir $l = 1$ setzen:

$$\begin{aligned} m_1(y_k) &= E(X \mid Y = y_k) = \sum_i x_i \frac{p_{ik}}{p_{\cdot k}}, \\ m_2(x_i) &= E(Y \mid X = x_i) = \sum_k y_k \frac{p_{ik}}{p_{i\cdot}}. \end{aligned} \tag{3.7.1}$$

Auf diese Weise erhalten wir zwei Punktmengen der x, y-Ebene. Die erste Menge besteht aus Punkten mit den Koordinaten

$$x = m_1(y_k), \quad y = y_k, \tag{3.7.2}$$

die zweite aus Punkten mit den Koordinaten

$$x = x_i, \quad y = m_2(x_i). \tag{3.7.3}$$

Nun sei (X, Y) eine zweidimensionale stetige Zufallsvariable mit der Dichte $f(x, y)$ und den Randdichten $f_1(x)$ und $f_2(y)$. Damit die bedingten Verteilungen der Veränderlichen X und Y definiert sind, nehmen wir an, daß die Dichtefunktionen den in 2.7 besprochenen zusätzlichen Bedingungen genügen. Die bedingten Mittelwerte $m_1(y)$ und $m_2(x)$ erhalten wir aus den Formeln (3.6.20) und (3.6.21):

$$\begin{aligned} m_1(y) &= E(X \mid Y = y) = \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_2(y)}\, dx, \\ m_2(x) &= E(Y \mid X = x) = \int_{-\infty}^{\infty} y \frac{f(x, y)}{f_1(x)}\, dy. \end{aligned} \tag{3.7.1'}$$

Wiederum ergeben sich zwei Punktmengen der x, y-Ebene, die eine mit den Koordinaten

$$x = m_1(y) \quad \text{und} \quad y, \tag{3.7.4}$$

die andere mit

$$x \quad \text{und} \quad y = m_2(x). \tag{3.7.5}$$

Definition 3.7.1. Die Punktmenge der x, y-Ebene mit den Koordinaten (3.7.2) oder (3.7.4) heißt *Regressionskurve der Zufallsvariablen X bezüglich Y*, die Punktmenge mit den Koordinaten (3.7.3) oder (3.7.5) heißt *Regressionskurve von Y bezüglich X*.

Aus der Definition folgt, daß auf der Regressionskurve von X bezüglich Y die bedingten Mittelwerte der Zufallsvariablen X liegen, wenn Y der Reihe nach alle möglichen Werte annimmt. Ebenso ist es bei der Regressionskurve von Y bezüglich X. Im allgemeinen überdecken sich die beiden Regressionskurven nicht.

Liegen alle Punkte der Regressionskurve auf einer Geraden, so sprechen wir von einer *geradlinigen Regression.*

Besteht zwischen den beiden Zufallsvariablen X und Y eine lineare Beziehung $Y = aX + b$, dann nimmt die Zufallsvariable (X, Y) nur Werte auf dieser Geraden an. Daraus folgt, daß die Mittelwerte auf der Geraden liegen und daß also auch beide Regressionskurven mit ihr zusammenfallen.

Sind die Zufallsvariablen X und Y unabhängig, so ist

$$m_2(x) = E(Y \mid X = x) = E(Y),$$

$$m_1(y) = E(X \mid Y = y) = E(X).$$

Der bedingte Mittelwert $m_2(x)$ hängt dann nicht von x ab. Die Regressionskurve der Zufallsvariablen Y bezüglich X liegt also auf einer zur x-Achse parallelen Geraden; ebenso liegt im Fall unabhängiger Zufallsvariabler die Regressionskurve von X bezüglich Y auf einer Parallelen zur y-Achse. Diese beiden Geraden schneiden sich in einem Punkt mit den Koordinaten (m_1, m_2).

Die hier besprochene Regression nennen wir im Unterschied zu einem anderen Begriff der Regression, den wir in 3.8 behandeln wollen, *Regression erster Art.*

Beispiel 3.7.1. Die Zufallsvariable (X, Y) kann die Wertepaare (x_k, y_l) $(k, l = 1, 2, 3, 4, 5)$ annehmen. Dabei ist

$$x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5, y_1 = 1, y_2 = 2, y_3 = 3, y_4 = 4, y_5 = 5;$$

die Wahrscheinlichkeiten p_{kl} der einzelnen Paare (x_k, y_l) sind in Tabelle 3.7.1 angegeben.

Außer den Werten p_{kl} ist in der letzten rechten Spalte dieser Tabelle die Randverteilung der Zufallsvariablen Y und in der untersten Zeile die Randverteilung der Zufallsvariablen X angegeben.

Tabelle 3.7.1. *Die Wahrscheinlichkeiten* p_{kl}

y_l \ x_k	1	2	3	4	5	Randverteilung von Y
1	$\frac{1}{12}$	$\frac{1}{24}$	—	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{1}{5}$
2	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{1}{5}$
3	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$	—	$\frac{1}{30}$	$\frac{1}{5}$
4	$\frac{1}{12}$	—	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{1}{5}$
5	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{1}{5}$
Randverteilung von X	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

In Tabelle 3.7.2 sind die bedingten Verteilungen der Zufallsvariablen Y unter der Bedingung $X = x_k$ $(k = 1, \ldots, 5)$ und in Tabelle 3.7.3 die bedingten Verteilungen der Zufallsvariablen X unter der Bedingung $Y = y_l$ $(l = 1, \ldots, 5)$ angegeben.

Tabelle 3.7.2.

y_l \ x_k	1	2	3	4	5
1	$\frac{1}{4}$	$\frac{1}{4}$	—	$\frac{1}{4}$	$\frac{1}{5}$
2	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{5}$
3	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	—	$\frac{1}{5}$
4	$\frac{1}{4}$	—	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{5}$
5	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{5}$
insgesamt	1	1	1	1	1

Tabelle 3.7.3.

y_l \ x_k	1	2	3	4	5	insgesamt
1	$\frac{5}{12}$	$\frac{5}{24}$	—	$\frac{5}{24}$	$\frac{1}{6}$	1
2	$\frac{5}{24}$	$\frac{5}{24}$	$\frac{5}{24}$	$\frac{5}{24}$	$\frac{1}{6}$	1
3	$\frac{5}{12}$	$\frac{5}{24}$	$\frac{5}{24}$	—	$\frac{1}{6}$	1
4	$\frac{5}{12}$	—	$\frac{5}{24}$	$\frac{5}{24}$	$\frac{1}{6}$	1
5	$\frac{5}{24}$	$\frac{5}{24}$	$\frac{5}{24}$	$\frac{5}{24}$	$\frac{1}{6}$	1

Wir berechnen die Mittelwerte der einen Zufallsvariablen, wenn die andere einen festen Wert annimmt. Es ist

$$E(Y\mid X=1)=\frac{1}{4}\cdot 1+\frac{1}{8}\cdot 2+\frac{1}{4}\cdot 3+\frac{1}{4}\cdot 4+\frac{1}{8}\cdot 5=2\,\frac{7}{8},$$

$$E(Y\mid X=2)=2\,\frac{3}{4},\quad E(Y\mid X=3)=3\,\frac{1}{2},\quad E(Y\mid X=4)=3,\quad E(Y\mid X=5)=3$$

und

$$E(X\mid Y=2)=\frac{5}{24}\cdot 1+\frac{5}{24}\cdot 2+\frac{5}{24}\cdot 3+\frac{5}{24}\cdot 4+\frac{1}{6}\cdot 5=2\,\frac{11}{12},$$

$$E(X\mid Y=1)=2\,\frac{1}{2},\quad E(X\mid Y=3)=2\,\frac{7}{24},\quad E(X\mid Y=4)=2\,\frac{17}{24},$$

$$E(X\mid Y=5)=2\,\frac{11}{12}.$$

Auf diese Weise erhalten wir zwei Punktmengen der x,y-Ebene. Die erste Menge besteht aus Punkten mit den Koordinaten

$$x=x_k,\qquad y=E(Y\mid X=x_k)\qquad (k=1,2,3,4,5)$$

und die zweite aus Punkten mit den Koordinaten

$$y=y_l,\qquad x=E(X\mid Y=y_l)\qquad (l=1,2,3,4,5).$$

Die erste Punktmenge ist die Regressionskurve der Zufallsvariablen Y bezüglich X, die zweite die Regressionskurve von X bezüglich Y.

Beispiel 3.7.2. Wir wollen die Regressionskurven für die in Beispiel 3.6.1 angegebene zweidimensionale Normalverteilung berechnen.

Wir bestimmen dazu zunächst die Randverteilungen

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_{-\infty}^{\infty}\frac{1}{2\pi}\,e^{-\frac{x^2-2xy+2y^2}{2}}\,dy=\frac{1}{2\sqrt{\pi}}\,e^{-\frac{x^2}{4}},$$

$$f_2(y)=\int_{-\infty}^{\infty} f(x,y)\,dx=\int_{-\infty}^{\infty}\frac{1}{2\pi}\,e^{-\frac{x^2-2xy+2y^2}{2}}\,dx=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{y^2}{2}}.$$

Nach (2.7.6) erhalten wir daraus die bedingten Dichten

$$f(y\mid x)=\frac{1}{\sqrt{\pi}}\,e^{-\left(y-\frac{x}{2}\right)^2},$$

$$f(x\mid y)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{(x-y)^2}{2}}$$

und hieraus

$$m_2(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} y e^{-\left(y - \frac{x}{2}\right)^2} dy = \frac{x}{2},$$

$$m_1(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-y)^2}{2}} dx = y.$$

Die Regressionskurven sind hier also Geraden. Später werden wir sehen, daß für jede Normalverteilung die Regressionskurven Geraden sind.

B. Die Regressionskurven erster Art haben eine äußerst wichtige Minimaleigenschaft. Die Regressionskurve von Y bezüglich X erfüllt nämlich die Relation

$$E\{[Y - m_2(X)]^2\} = \min. \tag{3.7.6}$$

Mit anderen Worten: Die mittlere quadratische Abweichung der Zufallsvariablen Y von der Funktion $u(X)$ ist am kleinsten, wenn $u(X)$ (mit der Wahrscheinlichkeit 1) gleich $m_2(X)$ ist.

Wir beweisen die Beziehung (3.7.6) für den Fall, daß (X, Y) eine zweidimensionale stetige Zufallsvariable mit der Dichte $f(x, y)$ ist. Dann ist

$$\begin{aligned} E\{[Y - u(X)]^2\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - u(x)]^2 f(x, y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} f_1(x) \int_{-\infty}^{\infty} [y - u(x)]^2 f(y \mid x)\, dy\, dx. \end{aligned} \tag{3.7.7}$$

Aus (3.2.12) folgt, daß die rechte Seite von (3.7.7) ihr Minimum für $u(x) = m_2(x)$ annimmt.

Für diskrete Zufallsvariable ist der Beweis analog.

Ebenso genügt die Regressionskurve von X bezüglich Y der Relation

$$E\{[X - m_1(Y)]^2\} = \min. \tag{3.7.8}$$

C. Für n-dimensionale $(n > 2)$ Zufallsvariable kann man den Begriff einer Regressionsfläche erster Art einführen. Wir beschränken uns auf stetige Zufallsvariable. Es sei also $f(x_1, x_2, \ldots, x_n)$ die Dichtefunktion der Zufallsvariablen $(X_1, X_2, \ldots, X_n)$. Das bedingte Moment, das durch die Formel (3.6.25) definiert wird, möge für $l = 1$ existieren. Wir wollen es mit $m_1(x_2, \ldots, x_n)$ bezeichnen. Es ist also

$$\begin{aligned} m_1(x_2, \ldots, x_n) &= E(X_1 \mid X_2 = x_2, \ldots, X_n = x_n) \\ &= \frac{\int_{-\infty}^{\infty} x_1 f(x_1, x_2, \ldots, x_n)\, dx_1}{\int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1}. \end{aligned}$$

Definition 3.7.2. Die Punktmenge des n-dimensionalen $(x_1, x_2, \ldots, x_n)$-Raumes mit den Koordinaten

$$x_1 = m_1(x_2, \ldots, x_n), \quad x_2, \quad \ldots, \quad x_n$$

heißt *Regressionsfläche erster Art der Zufallsvariablen* X_1 *bezüglich der Zufallsvariablen* $X_2, \ldots, X_n$.

Ebenso kann man auch die Regressionsflächen der übrigen Zufallsvariablen definieren.

Die Regressionsflächen erster Art besitzen eine analoge Minimaleigenschaft wie die Regressionskurven. Es gilt nämlich die Formel

$$E\{[X_1 - m_1(X_2, \ldots, X_n)]^2\} = \min, \tag{3.7.9}$$

d. h., die mittlere quadratische Abweichung der Zufallsvariablen X_1 von einer Funktion $u(X_2, \ldots, X_n)$ erreicht dann ihren kleinsten Wert, wenn die Funktion $u(X_2, \ldots, X_n)$ (mit der Wahrscheinlichkeit 1) gleich

$$m_1(X_2, \ldots, X_n)$$

ist.

3.8. Die Regression zweiter Art

A. Wir haben die Regressionskurve erster Art von Y bezüglich X als Punktmenge der x, y-Ebene mit den Koordinaten $x, m_2(x)$ definiert, wobei $m_2(x)$ der Mittelwert der Zufallsvariablen Y unter der Bedingung $X = x$ ist. Analog wurde die Regressionskurve der Zufallsvariablen X bezüglich Y definiert. Dabei zeigte es sich, daß die Regressionskurven den Bedingungen (3.7.6) bzw. (3.7.8) genügen. Diese Kurven liegen im allgemeinen nicht auf einer Geraden. Bei praktischen Problemen muß man aber manchmal diejenige Gerade bestimmen, von welcher (unter allen Geraden in der x, y-Ebene) die mittlere quadratische Abweichung der Zufallsvariablen Y am kleinsten ist. Hat also die Gleichung dieser Geraden die Form $y = \alpha x + \beta$, dann soll die Bedingung

$$E\{[Y - (\alpha X + \beta)]^2\} = \min \tag{3.8.1}$$

erfüllt sein.

Definition 3.8.1. Eine Gerade, die die in (3.8.1) geforderte Eigenschaft besitzt, heißt *Regressionsgerade zweiter Art der Zufallsvariablen* Y *bezüglich* X.

Wir wollen die Koeffizienten α und β dieser Geraden bestimmen. Wenn wir die Bezeichnungen

$$\begin{aligned} E(X) &= m_{10}, \quad E(Y) = m_{01}, \\ \sigma_1 &= \sqrt{\mu_{20}} = \sqrt{E[(X - m_{10})^2]}, \\ \sigma_2 &= \sqrt{\mu_{02}} = \sqrt{E[(Y - m_{01})^2]}, \\ \mu_{11} &= E[(X - m_{10})(Y - m_{01})] \end{aligned}$$

benutzen, so erhalten wir

$$\begin{aligned}
&E\{[Y-(\alpha X+\beta)]^2\}\\
&=E\{[Y-m_{01}-\alpha(X-m_{10})+m_{01}-\alpha m_{10}-\beta]^2\}\\
&=E[(Y-m_{01})^2]+\alpha^2E[(X-m_{10})^2]\\
&\quad-2\alpha E[(X-m_{10})(Y-m_{01})]+(m_{01}-\alpha m_{10}-\beta)^2\\
&=\mu_{02}+\alpha^2\mu_{20}-2\alpha\mu_{11}+(m_{01}-\alpha m_{10}-\beta)^2. \qquad (3.8.2)
\end{aligned}$$

Um das Minimum des Ausdrucks (3.8.2) zu bestimmen, differenzieren wir ihn nach α und nach β und setzen die Ableitungen gleich Null. Dann erhalten wir die beiden Gleichungen

$$\begin{aligned}
&2\alpha\mu_{20}-2\mu_{11}-2m_{10}(m_{01}-\alpha m_{10}-\beta)=0,\\
&m_{01}-\alpha m_{10}-\beta=0.
\end{aligned} \qquad (3.8.3)$$

Die Auflösung dieses Gleichungssystems ergibt

$$\alpha=\frac{\mu_{11}}{\mu_{20}}, \quad \beta=m_{01}-\alpha m_{10}.$$

Wenn wir die Bezeichnung $\varrho=\frac{\mu_{11}}{\sigma_1\sigma_2}$ für den Korrelationskoeffizienten benutzen, erhalten wir

$$\alpha=\varrho\frac{\sigma_2}{\sigma_1}, \quad \beta=m_{01}-\varrho\frac{\sigma_2}{\sigma_1}m_{10}. \qquad (3.8.4)$$

Die Koeffizienten α bzw. β werden wir in Zukunft mit α_{21} bzw. β_{21} bezeichnen.

Analog erhält man für die Koeffizienten der Regressionsgeraden von X bezüglich Y

$$\alpha_{12}=\varrho\frac{\sigma_1}{\sigma_2}, \quad \beta_{12}=m_{10}-\varrho\frac{\sigma_1}{\sigma_2}m_{01}. \qquad (3.8.5)$$

Die Größe α_{21} heißt *Regressionskoeffizient der Zufallsvariablen Y bezüglich X* und die Größe α_{12} *Regressionskoeffizient von X bezüglich Y*.

Die Gleichung der Regressionsgeraden von Y bezüglich X kann man also in der Form

$$y-m_{01}=\varrho\frac{\sigma_2}{\sigma_1}(x-m_{10}) \qquad (3.8.6)$$

schreiben, während die Gleichung der Regressionsgeraden von X bezüglich Y die Gestalt

$$y-m_{01}=\frac{1}{\varrho}\frac{\sigma_2}{\sigma_1}(x-m_{10}) \qquad (3.8.7)$$

hat. Aus den Gleichungen (3.8.6) und (3.8.7) folgt, daß im Fall $\varrho^2 = 1$ die Regressionsgeraden zusammenfallen.

Leicht lassen sich die Beziehungen

$$\begin{aligned} &E(Y - \alpha_{21}X - \beta_{21}) = 0, \\ &E\{[Y - (\alpha_{21}X + \beta_{21})]^2\} = \sigma_2^2(1 - \varrho^2) \end{aligned} \tag{3.8.8}$$

herleiten. Die Größe $\sigma_2^2(1 - \varrho^2)$ ist die Dispersion der Zufallsvariablen

$$Y - \alpha_{21}X - \beta_{21}$$

und gleichzeitig die minimale mittlere quadratische Abweichung der Zufallsvariablen Y von einer Geraden der x, y-Ebene.

Eine einfache Rechnung führt zu der Gleichung

$$E[(X - m_{10})(Y - \alpha_{21}X - \beta_{21})] = 0; \tag{3.8.9}$$

also sind der „Rest"

$$Y - \alpha_{21}X - \beta_{21}$$

und die Zufallsvariable X nicht korreliert.

Ebenso gelten die Beziehungen

$$\begin{aligned} &E(X - \alpha_{12}X - \beta_{12}) = 0, \\ &E\{[X - (\alpha_{12}Y + \beta_{12})]^2\} = \sigma_1^2(1 - \varrho^2), \\ &E[(Y - m_{01})(X - \alpha_{12}Y - \beta_{12})] = 0. \end{aligned} \tag{3.8.10}$$

Die Größen $\sigma_1^2(1 - \varrho^2)$ und $\sigma_2^2(1 - \varrho^2)$ nennt man *Restdispersionen.*

Beispiel 3.8.1. Wir wollen die Regressionsgeraden zweiter Art für das Beispiel 3.7.1 bestimmen. Zu diesem Zweck berechnen wir aus Tabelle 3.7.1 die Größen

$$m_{10} = E(X) = 2\tfrac{2}{3}, \qquad m_{01} = E(Y) = 3,$$

$$\sigma_1 = \sqrt{D^2(X)} = 1{,}49, \qquad \sigma_2 = \sqrt{D^2(Y)} = 1{,}41, \qquad \varrho = 0{,}06.$$

Die Gleichung der Regressionsgeraden von Y bezüglich X lautet hier wegen (3.8.6)

$$y - 3 = 0{,}056(x - 2\tfrac{2}{3}). \tag{3.8.11}$$

Die Regressionsgerade von X bezüglich Y hat wegen (3.8.7) die Gleichung

$$y - 3 = 15{,}7(x - 2\tfrac{2}{3}).$$

In Abb. 3.8.1 sind die durch die Formel (3.8.11) bestimmten Regressionsgeraden und die Regressionskurve von Y bezüglich X und in Abb. 3.8.2 die Regressionsgerade und die Regressionskurve von X bezüglich Y dargestellt.

Beispiel 3.8.2. Es sind die Gleichungen der Regressionsgeraden für die zweidimensionale Normalverteilung des Beispiels 3.6.1 zu bestimmen.

Für dieses Beispiel hatten wir schon die Werte

$$m_{01} = m_{10} = 0, \qquad \sigma_1 = \sqrt{2}, \qquad \sigma_2 = 1, \qquad \varrho = \frac{1}{\sqrt{2}}$$

berechnet. Setzen wir sie in die Formeln (3.8.4) und (3.8.5) ein, so erhalten wir

$$\alpha_{21} = \frac{1}{2}, \qquad \beta_{21} = 0, \qquad \alpha_{12} = 1, \qquad \beta_{12} = 0.$$

Die Gleichungen der Regressionsgeraden lauten also

$$y = \frac{1}{2}x,$$

$$y = x.$$

Wie wir sehen, sind hier die Regressionsgeraden zweiter Art mit den Regressionskurven erster Art identisch (vgl. Beispiel 3.7.2). Diese Eigenschaft besitzt jede zweidimensionale Normalverteilung.

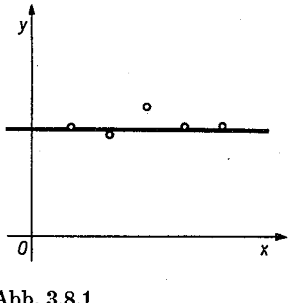


Abb. 3.8.1

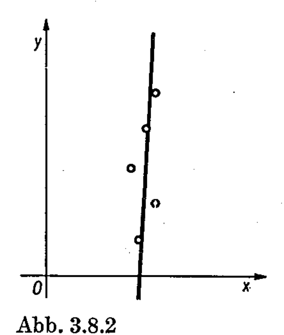


Abb. 3.8.2

B. Die Verallgemeinerung des Begriffs der Regressionsgeraden zweiter Art auf n-dimensionale Zufallsvariable $(n > 2)$ ist der Begriff der Regressionshyperebene zweiter Art.

Es sei $(X_1, X_2, \ldots, X_n)$ eine n-dimensionale Zufallsvariable und $E(X_i) = 0$ $(i = 1, 2, \ldots, n)$. Diese Voraussetzung bedeutet keine wesentliche Beschränkung der Allgemeinheit. Wäre sie nämlich nicht erfüllt, so könnten wir die Zufallsvariablen $X_i - E(X_i)$ betrachten. Wir haben gezeigt, daß die Regressionsfläche erster Art der Zufallsvariablen X_1 bezüglich der übrigen Veränderlichen der Bedingung (3.7.9) genügt. Nicht immer ist aber die Regressionsfläche erster Art eine Hyperebene.

Definition 3.8.2. Die Hyperebene

$$X_1 = \alpha_{12}X_2 + \alpha_{13}X_3 + \cdots + \alpha_{1n}X_n, \tag{3.8.12}$$

die der Bedingung

$$E[(X_1 - \alpha_{12}X_2 - \alpha_{13}X_3 - \cdots - \alpha_{1n}X_n)^2] = \min \tag{3.8.13}$$

genügt, nennt man *Regressionshyperebene zweiter Art der Zufallsvariablen* X_1 *bezüglich der Zufallsvariablen* $X_2, X_3, \ldots, X_n$.

Wir nehmen an, daß die Determinante der durch die Formel (3.6.30) definierten Matrix M von Null verschieden ist. Wäre das nicht der Fall, so würde nach Satz 3.6.6 zwischen den Zufallsvariablen $X_1, X_2, \ldots, X_n$ mindestens eine lineare Beziehung bestehen, und die durch diese lineare Beziehung bestimmte Hyperebene würde den Bedingungen (3.8.13) genügen.

Wir schreiben (3.8.13) in der Form

$$\sigma_1^2 + \sum_{i,k=2}^{n} \lambda_{ik}\alpha_{1i}\alpha_{1k} - 2\sum_{i=2}^{n} \lambda_{1i}\alpha_{1i} = \min,$$

differenzieren diesen Ausdruck nach den α_{1i} und erhalten das lineare Gleichungssystem

$$\begin{aligned}
\lambda_{22}\alpha_{12} + \lambda_{23}\alpha_{13} + \cdots + \lambda_{2n}\alpha_{1n} &= \lambda_{12},\\
\lambda_{32}\alpha_{12} + \lambda_{33}\alpha_{13} + \cdots + \lambda_{3n}\alpha_{1n} &= \lambda_{13},\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\\
\lambda_{n2}\alpha_{12} + \lambda_{n3}\alpha_{13} + \cdots + \lambda_{nn}\alpha_{1n} &= \lambda_{1n}.
\end{aligned} \tag{3.8.14}$$

Nach Voraussetzung ist die Determinante $|M| \neq 0$; die symmetrische quadratische Form (3.6.31) ist folglich positiv definit. Daraus folgt, daß alle Hauptminoren der Determinante $|M|$ positiv sind.[1]) Das Gleichungssystem (3.8.14) ist also eindeutig lösbar, und wir erhalten

$$\alpha_{1i} = -\frac{|M_{1i}|}{|M_{11}|} \qquad (i = 2, \ldots, n), \tag{3.8.15}$$

wobei $|M_{1i}|$ und $|M_{11}|$ die zu λ_{1i} bzw. λ_{11} gehörigen Minoren der Determinante $|M|$ sind.

Wir setzen nun

$$Y_1 = X_1 - \alpha_{12}X_2 - \cdots - \alpha_{1n}X_n.$$

Wenn wir für $\alpha_{12}, \ldots, \alpha_{1n}$ die Ausdrücke (3.8.15) einsetzen, erhalten wir

$$Y_1 = X_1 + \frac{1}{|M_{11}|}\sum_{i=2}^{n} |M_{1i}|X_i. \tag{3.8.16}$$

---

[1]) Siehe Gantmacher [1], S. 281, oder Smirnow [1], S. 112.

Es ist $E(Y_1) = 0$. Weiter erhalten wir

$$E(Y_1 X_k) = E\left[X_1 X_k + \frac{1}{|M_{11}|}\sum_{i=2}^{n} |M_{1i}| X_i X_k\right]$$

$$= \lambda_{1k} + \frac{1}{|M_{11}|}\sum_{i=2}^{n} \lambda_{ik}|M_{1i}| = \frac{1}{|M_{11}|}\sum_{i=1}^{n} \lambda_{ik}|M_{1i}|.$$

Daraus folgt

$$E(Y_1 X_k) = \begin{cases} \dfrac{|M|}{|M_{11}|} & \text{für} \quad k = 1, \\ 0 & \text{für} \quad k = 2, \ldots, n. \end{cases} \tag{3.8.17}$$

Also ergibt sich, daß der „Rest“ Y_1 mit keiner der Zufallsvariablen $X_2, \ldots, X_n$ korreliert ist.

Wir erhalten weiter unter Berücksichtigung von (3.8.17)

$$D^2(Y_1) = E(Y_1^2) = E(Y_1 X_1) = \frac{|M|}{|M_{11}|}. \tag{3.8.18}$$

Der Ausdruck (3.8.18) heißt *Restdispersion.*

Analog erhalten wir für die Koeffizienten der Regressionshyperebene von X_k bezüglich der übrigen Zufallsvariablen die Formel

$$\alpha_{ki} = -\frac{|M_{ki}|}{|M_{kk}|} \quad (i = 1, \ldots, k-1, \ k+1, \ldots, n). \tag{3.8.19}$$

Der „Rest“ wird durch die Formel

$$Y_k = X_k + \frac{1}{|M_{kk}|}\sum_{\substack{i=1 \\ i \neq k}}^{n} |M_{ki}| X_i \tag{3.8.20}$$

gegeben. Die Zufallsvariable Y_k ist mit keiner der Zufallsvariablen X_i $(i \neq k)$ korreliert, und die Restdispersion ist gleich

$$D^2(Y_k) = \frac{|M|}{|M_{kk}|}. \tag{3.8.21}$$

Beispiel 3.8.3. Die Zufallsvariable (X_1, X_2, X_3) habe eine dreidimensionale Normalverteilung mit der Dichtefunktion

$$f(x_1, x_2, x_3) = \frac{\exp\left\{-\frac{1}{4}\left[3x_1^2 + 3x_2^2 + 3x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3\right]\right\}}{(2\pi)^{\frac{3}{2}}}.$$

Durch einfache Rechnungen erhält man die Momentenmatrix

$$M = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}.$$

Wir wollen die Regressionsebene zweiter Art von X_1 bezüglich X_2 und X_3 bestimmen. Aus der Formel (3.8.15) erhalten wir

$$\alpha_{12} = \frac{\begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{vmatrix}} = \frac{1}{3}, \qquad \alpha_{13} = -\frac{\begin{vmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix}}{\begin{vmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{vmatrix}} = \frac{1}{3}.$$

Die Gleichung der Regressionshyperebene der Zufallsvariablen X_1 hat also die Form

$$X_1 = \frac{1}{3} X_2 + \frac{1}{3} X_3.$$

Nach (3.8.18) ist die Dispersion der Zufallsvariablen

$$Y_1 = X_1 - \frac{1}{3} X_2 - \frac{1}{3} X_3$$

gleich $\frac{2}{3}$. Weiter bemerken wir, daß die Dispersion der Zufallsvariablen X_1 gleich 1 ist.

Wie der Leser nachprüfen möge, fällt die Regressionsfläche erster Art der Zufallsvariablen X_1 mit der Regressionshyperebene zweiter Art zusammen. Diese Eigenschaft besitzen alle mehrdimensionalen Normalverteilungen.

C. Das Verfahren zur Bestimmung der unbekannten Koeffizienten der Geraden und Hyperebene, das wir in 3.8 angewandt haben, heißt die *Methode der kleinsten Quadrate.* Die Grundgedanken dieses Verfahrens stammen von Legendre [1] und Gauss [1]. Die weitere Entwicklung dieser Ideen, die für die Wahrscheinlichkeitsrechnung und mathematische Statistik äußerst wichtig sind, verdanken wir Markoff [5]. Wir empfehlen dem Leser auch die Arbeit von David und Neyman [1]. In dem Buch von Linnik [3] findet der Leser die Beschreibung der wichtigsten Ergebnisse, die bisher auf dem Gebiet der Methode der kleinsten Quadrate erzielt wurden.

3.9. Aufgaben und Ergänzungen

1. Man drücke das zentrale Moment μ_4 als Funktion der gewöhnlichen Momente m_1, m_2, m_3 und m_4 aus.
2. Man drücke das gewöhnliche Moment m_4 als Funktion der zentralen Momente μ_2, μ_3 und μ_4 aus.
3. Man bestimme für die Zufallsvariable X aus dem Beispiel 3.2.3 die Momente m_4 und μ_4.
4. Es seien X_1 und X_2 unabhängig, und sie mögen die gleiche Verteilung aufweisen. Man zeige, daß dann die Zufallsvariable $Y = X_1 - X_2$ eine symmetrische Verteilung hat.

Die Ergebnisse der nun folgenden Aufgaben 5 bis 8 stammen von Kolmogoroff [7].

5. Die reelle Funktion $g(x)$ sei nichtnegativ. Sie genügt daher der Ungleichung $g(x) \geqq b > 0$ für $x \geqq a$. Man zeige, daß für jede Zufallsvariable X, für die $E\big(g(X)\big)$ existiert, die folgende Beziehung gilt:
$$P(X \geqq a) \leqq \frac{E\big(g(X)\big)}{b}.$$
6. a) Es sei $g(x)$ eine nichtnegative, gerade und außerdem für $x > 0$ nichtfallende Funktion. Man zeige, daß für jede Zufallsvariable X, für die $E\big(g(X)\big)$ existiert, die Beziehung
$$P(|X| \geqq \varepsilon) \leqq \frac{E\big(g(X)\big)}{g(\varepsilon)}$$
für jedes $\varepsilon > 0$ gilt.

   b) Man beweise mit Hilfe dieser Ungleichung die Tschebyscheffsche Ungleichung.
7. Eine reelle Funktion $g(x)$ möge für alle x der Ungleichung $|g(x)| \leqq K$ genügen. Man zeige, daß die Beziehung
$$P(|X| \geqq \varepsilon) \geqq \frac{E\big(g(X)\big) - g(\varepsilon)}{K}$$
für jedes $\varepsilon > 0$ gilt.
8. Es sei $P(|X| \leqq M) = 1$, und es möge die reelle Funktion $g(x)$ allen Bedingungen der Aufgabe 6a genügen. Man zeige, daß für jedes $\varepsilon > 0$
$$P(|X| \geqq \varepsilon) \geqq \frac{E\big(g(X)\big) - g(\varepsilon)}{g(M)}$$
und insbesondere
$$P(|X| \geqq \varepsilon) \geqq \frac{E(X^2) - \varepsilon^2}{M^2}$$
gilt.
9. Man zeige (Berge [1]), daß für beliebige Zufallsvariable Y_1 und Y_2 mit den Standardabweichungen σ_1 und σ_2 und dem Korrelationskoeffizienten ϱ für jedes k die Ungleichung
$$P(|Y_1 - E(Y_1)| \geqq k\sigma_1 \text{ oder } |Y_2 - E(Y_2)| \geqq k\sigma_2) \leqq \frac{1 + \sqrt{1 - \varrho^2}}{k^2}$$
erfüllt ist. (Eine Verallgemeinerung für Zufallsvektoren mit der Dimension > 2 haben Olkin und Pratt [1] angegeben.)

10. Es sei ϱ der Korrelationskoeffizient der Zufallsvariablen X_1 und X_2. Man zeige, daß dann ϱ auch der Korrelationskoeffizient der Variablen $Y_1 = a_1 X_1 + b_1$ und $Y_2 = a_2 X_2 + b_2$ ist.
11. Die beiden Zufallsvariablen X und Y mögen den gleichen Erwartungswert 0 und gleiche Varianz haben. Dann sind die Zufallsvariablen

$$Z = X + Y, \qquad U = X - Y$$

nicht korreliert. Sind sie notwendigerweise auch unabhängig? (Man vergleiche hierzu den Satz von Skitowitsch in 5.7.)
12. Die Zufallsvariablen X und Y mögen endliche Momente zweiter Ordnung besitzen; Z und U seien durch die linearen Funktionen

$$Z = aX + bY + c,$$
$$U = dX + eY + f$$

definiert. Man bestimme die Koeffizienten a, b, c, d, e, f so, daß Z und Y nicht korreliert sind.
13. Der Ausdruck

$$\Theta_{YX} = \frac{1}{D^2(Y)} E\big(m_2(X) - E(Y)\big)^2,$$

in dem $m_2(X)$ durch (3.7.1) oder (3.7.1') gegeben ist, heißt das *Korrelationsverhältnis* der Zufallsvariablen Y bezüglich der Zufallsvariablen X.

a) Man beweise folgende Behauptungen:

I. $0 \leqq \Theta_{YX}^2 \leqq 1$.

II. Die Gleichung $\Theta_{YX}^2 = 0$ gilt genau dann, wenn $m_2(X) \equiv \text{const}$ ist.

III. Die Gleichung $\Theta_{YX}^2 = 1$ gilt genau dann, wenn $P(Y - m_2(X) = 0) = 1$ gilt.

b) Wofür ist das Korrelationsverhältnis ein Maß?
14. Man berechne die Korrelationsverhältnisse von Y bezüglich X und von X bezüglich Y im Beispiel 2.6.1 und in den Beispielen 3.6.1 und 3.7.1.
15. Es sei

$$Y_1 = X_1 - a_{13} X_3 - \cdots - a_{1n} X_n,$$
$$Y_2 = X_2 - a_{23} X_3 - \cdots - a_{2n} X_n,$$

wobei $E(X_i) = 0$ $(i = 1, 2, \ldots, n)$ ist und die Koeffizienten a_{1i} und a_{2i} $(i = 3, \ldots, n)$ aus Formeln bestimmt wurden, die der Formel (3.8.15) analog sind. Der Ausdruck

$$\varrho_{12\cdot 3\ldots n} = \frac{E(Y_1 Y_2)}{\sqrt{E(Y_1^2)\, E(Y_2^2)}}$$

heißt der *partielle Korrelationskoeffizient der Variablen X_1 und X_2 bezüglich der Variablen $X_3, \ldots, X_n$*. Er ist ein Maß für die Korrelation der Zufallsvariablen X_1 und X_2 nach Elimination des Einflusses der Zufallsvariablen $X_3, \ldots, X_n$.

a) Man beweise (bezüglich der Bezeichnungen vgl. 3.8.)

$$\varrho_{12\cdot 3\ldots n} = -\frac{|M_{12}|}{\sqrt{|M_{11}|\,|M_{12}|}}.$$

b) Man beweise

$$\varrho_{12\cdot3} = \frac{\varrho_{12} - \varrho_{13}\varrho_{23}}{\sqrt{(1 - \varrho_{23}^2)(1 - \varrho_{13}^2)}},$$

wobei ϱ_{12}, ϱ_{13} bzw. ϱ_{23} die gewöhnlichen Korrelationskoeffizienten für die Paare der Zufallsvariablen (X_1, X_2), (X_1, X_3) bzw. (X_2, X_3) sind.

16. Man ermittle sämtliche partiellen Korrelationskoeffizienten für das Beispiel 3.8.3.

17. Es sei

$$X_1^* = \alpha_{12}X_2 + \cdots + \alpha_{1n}X_n,$$

wobei $E(X_i) = 0$ $(i = 1, \ldots, n)$ ist und die α_{1i} $(i = 2, \ldots, n)$ durch (3.8.15) gegeben sind. Der Korrelationskoeffizient der Zufallsvariablen X_1^* und X_1 heißt der *multiple Korrelationskoeffizient zwischen* X_1 *und* $(X_2, \ldots, X_n)$. Wir bezeichnen ihn mit $\varrho_{1(2\ldots n)}$; es gilt dann

$$\varrho_{1(2\ldots n)} = \frac{E(X_1X_1^*)}{\sqrt{E(X_1^2)\,E(X_1^{*2})}}.$$

Man zeige, daß unter allen Linearkombinationen der Zufallsvariablen $X_2, \ldots, X_n$ die Zufallsvariable X_1^* am stärksten mit X_1 korreliert ist.

18. Man berechne sämtliche multiplen Korrelationskoeffizienten im Beispiel 3.8.3.

19. Man beweise: Die Wahrscheinlichkeit dafür, daß zwischen den Zufallsvariablen $X_1, \ldots, X_n$ mit endlichen Varianzen und Kovarianzen genau $n - r$ lineare Beziehungen bestehen, ist dann und nur dann gleich 1, wenn der Rang der Matrix M der Momente zweiter Ordnung gleich r ist. Wir sagen dann, die Verteilung sei vom *Rang* r (Frisch [1]).

20. Es sei $\{X_n\}$ $(n = 1, 2, 3, \ldots)$ eine Folge von Zufallsvariablen mit den Erwartungswerten $E(X_n)$. Man zeige: Ist

$$\sum_{n=1}^{\infty} E(|X_n|) < \infty,$$

so gilt

$$E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n).$$

Man folgere hieraus, daß die Reihe $\sum_{n=1}^{\infty} X_n$ mit der Wahrscheinlichkeit 1 konvergent ist.

Hinweis. Man stütze sich auf den Satz über die monotone Konvergenz der Folge $\{Y_n\}$, wobei $Y_n = \sum_{k=1}^{n} |X_k|$ ist.

Siehe auch Aufgabe 4.8.10.

4. CHARAKTERISTISCHE FUNKTIONEN

4.1. Die Eigenschaften der charakteristischen Funktionen

In diesem Kapitel untersuchen wir die Mittelwerte einer gewissen Funktion einer Zufallsvariablen und erhalten damit ein Hilfsmittel, das sich bei den weiteren Untersuchungen der Wahrscheinlichkeitsrechnung und ihrer Anwendungen auf die Statistik als sehr nützlich erweisen wird.

Definition 4.1.1. Als *charakteristische Funktion* der Zufallsvariablen X bezeichnen wir die Funktion

$$\varphi(t) = E(e^{itX}), \tag{4.1.1}$$

wobei t eine reelle Zahl und i die imaginäre Einheit ist.

Ist X eine diskrete Zufallsvariable mit den Sprungstellen

$$x_k \ (k = 1, 2, \ldots)$$

und $P(X = x_k) = p_k$, dann wird die charakteristische Funktion von X durch

$$\varphi(t) = E(e^{itX}) = \sum_k p_k e^{itx_k} \tag{4.1.2}$$

gegeben. Wegen

$$|e^{itx_k}| = 1 \quad \text{und} \quad \sum_k p_k = 1$$

ist die Reihe auf der rechten Seite der Formel (4.1.2) absolut und gleichmäßig konvergent. Die charakteristische Funktion $\varphi(t)$ als Summe einer gleichmäßig konvergenten Reihe von stetigen Funktionen ist also für jeden reellen Wert t stetig.

Beispiel 4.1.1. Die Zufallsvariable X soll die Werte $x_1 = -1$ und $x_2 = 1$ mit den Wahrscheinlichkeiten

$$P(X = -1) = P(X = 1) = 0{,}5$$

annehmen können. Es ist die charakteristische Funktion dieser Zufallsvariablen zu bestimmen.

Durch Anwendung der Formel (4.1.2) erhalten wir

$$\varphi(t) = 0{,}5e^{-it} + 0{,}5e^{it} = 0{,}5(\cos t - i\sin t) + 0{,}5(\cos t + i\sin t) = \cos t. \tag{4.1.3}$$

Ist X eine stetige Zufallsvariable mit der Dichte $f(x)$, dann wird die charakteristische Funktion dieser Zufallsvariablen durch

$$\varphi(t) = E(e^{itX}) = \int_{-\infty}^{\infty} f(x) e^{itx}\, dx \tag{4.1.4}$$

gegeben. Da

$$\int_{-\infty}^{\infty} f(x) |e^{itx}|\, dx = \int_{-\infty}^{\infty} f(x)\, dx = 1$$

ist, konvergiert das Integral in der Formel (4.1.4) absolut und gleichmäßig; $\varphi(t)$ ist also eine stetige Funktion von t.

Beispiel 4.1.2. Die Dichte $f(x)$ sei folgendermaßen erklärt:

$$f(x) = \begin{cases} 0 & \text{für} \quad x < 0, \\ 1 & \text{für} \quad 0 \leqq x \leqq 1, \\ 0 & \text{für} \quad x > 1. \end{cases} \tag{4.1.5}$$

Wir sagen, die Zufallsvariable sei im Intervall $[0, 1]$ *gleichverteilt*. Ihre charakteristische Funktion ist

$$\varphi(t) = \int_{-\infty}^{\infty} f(x)\, e^{itx}\, dx = \int_{0}^{1} f(x)\, e^{itx} dx = \left[\frac{e^{itx}}{it}\right]_0^1 = \frac{e^{it} - 1}{it}. \tag{4.1.6}$$

Wir untersuchen einige weitere Eigenschaften der charakteristischen Funktionen. Es ist

$$\varphi(0) = E(e^0) = E(1) = 1. \tag{4.1.7}$$

Weiter erhalten wir

$$|\varphi(t)| = |E(e^{itX})| \leqq E(|e^{itX}|) = 1,$$

also

$$|\varphi(t)| \leqq 1. \tag{4.1.8}$$

Wegen

$$\varphi(-t) = E(e^{-itX}) = E(\cos t X - i \sin t X) = E(\cos t X) - iE(\sin t X)$$

und

$$\varphi(t) = E(e^{itX}) = E(\cos t X + i \sin t X) = E(\cos t X) + iE(\sin tX)$$

erhalten wir

$$\varphi(-t) = \overline{\varphi(t)}, \tag{4.1.9}$$

wobei $\overline{\varphi(t)}$ die zu $\varphi(t)$ konjugiert komplexe Zahl ist.

Jede charakteristische Funktion muß die Bedingungen (4.1.7), (4.1.8) und (4.1.9) erfüllen; diese sind jedoch nicht hinreichend. Nicht jede Funktion $\varphi(t)$, die diesen Bedingungen genügt, ist die charakteristische Funktion einer Zufallsvariablen.

Eine andere notwendige Bedingung hat MARCINKIEWICZ [1] angegeben. Er hat gezeigt, daß eine nicht identisch konstante Funktion $\varphi(t)$, die sich in einer Umgebung von $t = 0$ in der Form

$$\varphi(t) = 1 + O(t^{2+\alpha})$$

darstellen läßt ($\alpha > 0$), keine charakteristische Funktion sein kann. Hieraus folgt z. B. sofort, daß weder $\varphi(t) = \exp(-t^4)$ noch $\varphi(t) = 1/(1 + t^4)$ charakteristische Funktionen sind. Weitere notwendige Bedingungen hat LUKACS [4] angegeben. Wir wollen hier ohne Beweis den Satz von BOCHNER [2] anführen, der die notwendigen und hinreichenden Bedingungen dafür angibt, daß eine Funktion $\varphi(t)$ der reellen Variablen t eine charakteristische Funktion ist.

Satz 4.1.1. *Eine für* $-\infty < t < \infty$ *definierte Funktion* $\varphi(t)$ *möge der Gleichung* (4.1.7) *genügen. Die Funktion* $\varphi(t)$ *ist dann und nur dann eine charakteristische Funktion, wenn sie*

1. *stetig ist;*

2. *für* $n = 1, 2, \ldots$ *und für beliebige reelle* $t_1, \ldots, t_n$ *sowie für komplexe* $a_1, \ldots, a_n$ *der Ungleichung*

$$\sum_{j,k=1}^{n} \varphi(t_j - t_k) a_j \bar{a}_k \geqq 0$$

genügt.

Den verhältnismäßig einfachen Beweis dieses Satzes findet der Leser in der Monographie von LOÈVE [4].

Wir erinnern daran, daß eine Funktion, die der Bedingung 2 des Satzes 4.1.1 genügt, *positiv definit* genannt wird.

Eine andere notwendige und hinreichende Bedingung dafür, daß eine Funktion $\varphi(t)$ eine charakteristische Funktion ist, hat CRAMÉR [5] angegeben (siehe Aufgabe 4.8.3).

4.2. Charakteristische Funktionen und Momente

A. Wir betrachten die Zufallsvariable X und setzen voraus, daß ihr Moment $m_l = E(X^l)$ der Ordnung l existiert.

Es sei X eine diskrete Zufallsvariable mit den Sprungstellen x_k. Dann kann man die Formel (4.1.2) l-mal nach t differenzieren.

In der Tat, die l-te Ableitung des Ausdrucks unter dem Summenzeichen in der Formel (4.1.2) nach t ist gleich

$$p_k i^l x_k^l e^{itx_k}.$$

Aus der Existenz des Moments l-ter Ordnung folgt die Existenz des absoluten Moments l-ter Ordnung. Da

$$\sum_k |i^l p_k x_k^l e^{itx_k}| = \sum_k |p_k x_k^l| = \beta_l$$

ist, darf man in der Formel für die charakteristische Funktion l-mal unter dem Summenzeichen differenzieren. Wir erhalten also

$$\varphi^{(l)}(t) = \sum_k p_k i^l x_k^l e^{itx_k} = E[i^l X^l e^{itX}]. \tag{4.2.1}$$

Nun sei $f(x)$ die Dichte einer stetigen Zufallsvariablen X. Dann kann man die Formel (4.1.4) l-mal nach t differenzieren.

In der Tat, die l-te Ableitung des Integranden in der Formel (4.1.4) nach t ist gleich

$$i^l x^l f(x) e^{itx}.$$

Weiter erhalten wir

$$\int_{-\infty}^{\infty} |i^l x^l f(x) e^{itx}|\, dx = \int_{-\infty}^{\infty} |x^l f(x)|\, dx = \beta_l.$$

Nach Voraussetzung ist das absolute Moment β_l endlich; folglich darf man in der Formel für die Funktion $\varphi(t)$ l-mal unter dem Integralzeichen differenzieren. Somit erhalten wir

$$\varphi^{(l)}(t) = \int_{-\infty}^{\infty} i^l x^l f(x) e^{itx}\, dx = E[i^l X^l e^{itX}]. \tag{4.2.2}$$

Wir haben damit für eine stetige Zufallsvariable dasselbe Ergebnis bekommen wie für eine diskrete Zufallsvariable:

$$\varphi^{(l)}(t) = E[i^l X^l e^{itX}]. \tag{4.2.3}$$

Aus der Gleichung (4.2.3) berechnen wir $\varphi^{(l)}(0)$ und erhalten

$$\varphi^{(l)}(0) = i^l E(X^l) = i^l m_l,$$

also

$$m_l = \frac{\varphi^{(l)}(0)}{i^l}. \tag{4.2.4}$$

Somit haben wir den folgenden Satz bewiesen:

Satz 4.2.1. *Existiert das l-te Moment einer Zufallsvariablen X, dann wird es durch die Formel* (4.2.4) *gegeben, in der $\varphi^{(l)}(0)$ die l-te Ableitung der charakteristischen Funktion $\varphi(t)$ von X im Punkt $t = 0$ ist.*

Beispiel 4.2.1. Die Zufallsvariable X habe eine Poissonverteilung, d. h., sie kann die Werte $x_k = k$ annehmen, wobei k eine ganze nichtnegative Zeit ist. Ihre Wahrscheinlichkeitsfunktion wird durch

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{4.2.5}$$

gegeben; dabei ist λ eine gewisse positive Konstante. Wir wollen die charakteristische Funktion dieser Verteilung bestimmen.

Nach der Formel (4.1.2) erhalten wir

$$\varphi(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}. \tag{4.2.6}$$

Weiter ist

$$\varphi'(t) = \lambda i e^{it} e^{\lambda(e^{it}-1)}.$$

Auf Grund der Formel (4.2.4) erhalten wir

$$m_1 = \frac{\varphi'(0)}{i} = \frac{\lambda i}{i} = \lambda. \tag{4.2.7}$$

Ebenso berechnen wir

$$\varphi''(t) = \lambda i^2 e^{it} e^{\lambda(e^{it}-1)} (\lambda e^{it} + 1),$$

also

$$m_2 = \frac{\varphi''(0)}{i^2} = \frac{i^2 \lambda(\lambda+1)}{i^2} = \lambda(\lambda+1). \tag{4.2.8}$$

Das zentrale Moment zweiter Ordnung ist somit

$$\mu_2 = \lambda(\lambda+1) - \lambda^2 = \lambda. \tag{4.2.9}$$

In ähnlicher Weise kann man die Momente höherer Ordnung berechnen.

Beispiel 4.2.2. Wir wollen die charakteristische Funktion einer Normalverteilung und ihre Momente berechnen. Es ist

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

also

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}} e^{-\frac{t^2}{2}} dx = e^{-\frac{t^2}{2}}. \tag{4.2.10}$$

Da $\varphi'(t) = -te^{-\frac{t^2}{2}}$ ist, ergibt sich

$$m_1 = \frac{\varphi'(0)}{i} = 0. \tag{4.2.11}$$

Weiter erhalten wir

$$\varphi''(t) = e^{-\frac{t^2}{2}}(t^2 - 1),$$

also

$$m_2 = \frac{\varphi''(0)}{i^2} = 1. \tag{4.2.12}$$

Diese Werte für m_1 und m_2 haben wir schon in den Beispielen 3.1.4 und 3.2.2 erhalten. Wie der Leser nachprüfen kann, sind alle ungeraden Momente gleich Null, während die geraden Momente durch die Formel

$$m_{2l} = 1 \cdot 3 \cdot 5 \cdots (2l - 1) \tag{4.2.13}$$

gegeben werden.

Wir weisen darauf hin, daß die Umkehrung des Satzes 4.2.1 nicht gilt. In Aufgabe 4.8.9 findet der Leser ein Beispiel für eine Zufallsvariable, deren Erwartungswert nicht existiert und deren charakteristische Funktion an der Stelle $t = 0$ differenzierbar ist. Besitzt jedoch die charakteristische Funktion $\varphi(t)$ an der Stelle $t = 0$ eine endliche Ableitung der geraden Ordnung $2k$, dann existiert ein Moment der Ordnung $2k$ der entsprechenden Zufallsvariablen (vgl. Aufgabe 4.8.8). Dann existieren aber bekanntlich auch alle Momente, deren Ordnung kleiner ist als $2k$.

B. Wir wollen nun untersuchen, wie sich die charakteristische Funktion der Zufallsvariablen X verhält, wenn wir X linear transformieren.

Wir betrachten zunächst die Translation

$$Y = X + b.$$

Wenn wir mit $\varphi_1(t)$ die charakteristische Funktion der Zufallsvariablen Y bezeichnen, erhalten wir

$$\varphi_1(t) = E(e^{itY}) = E(e^{it(X+b)}) = E(e^{itX})\, e^{itb} = e^{itb}\varphi(t). \tag{4.2.14}$$

Wir sehen also, daß bei einer Translation um eine Konstante b die charakteristische Funktion mit dem Faktor e^{itb} multipliziert wird.

Es sei jetzt

$$Y = aX.$$

Wir erhalten

$$\varphi_1(t) = E(e^{itY}) = E(e^{itaX}) = \varphi(at). \tag{4.2.15}$$

Die charakteristische Funktion der Zufallsvariablen aX ist also gleich der charakteristischen Funktion der Zufallsvariablen X im Punkt at.

Insbesondere ergibt sich für $a = -1$

$$\varphi_1(t) = \varphi(-t) = \overline{\varphi(t)}.$$

Wir betrachten nun die Transformation

$$Y = aX + b.$$

Bezeichnen wir mit $\varphi(t)$ bzw. $\varphi_1(t)$ die charakteristischen Funktionen der Zufallsvariablen X bzw. Y, so erhalten wir aus den Gleichungen (4.2.14) und (4.2.15)

$$\varphi_1(t) = e^{ibt}\varphi(at). \tag{4.2.16}$$

Insbesondere sei

$$Y = \frac{X - m_1}{\sigma},$$

wobei m_1 bzw. σ der Mittelwert bzw. die Standardabweichung der Zufallsvariablen X sind. Wir erhalten dann

$$\varphi_1(t) = e^{-\frac{m_1 it}{\sigma}}\varphi\left(\frac{t}{\sigma}\right). \tag{4.2.17}$$

4.3. Die Semiinvarianten

Manchmal ist es angebracht, die Verteilung einer Zufallsvariablen nicht durch die Momente, sondern durch eine andere Gruppe von Parametern zu charakterisieren, die wir erhalten, wenn wir die Funktion

$$\psi(t) = \log\varphi(t) \tag{4.3.1}$$

betrachten; dabei ist $\varphi(t)$ die charakteristische Funktion der Zufallsvariablen.

Wir entwickeln die Funktion $\varphi(t)$ in der Umgebung des Punktes $t = 0$ formal in eine unendliche Reihe:

$$\varphi(t) = 1 + \sum_{s=1}^{\infty} \frac{m_s}{s!}(it)^s. \tag{4.3.2}$$

Mit z bezeichnen wir die um 1 verminderte rechte Seite von (4.3.2) und entwickeln die Funktion $\psi(t)$ formal in eine Potenzreihe:

$$\psi(t) = \log\varphi(t) = \log(1 + z) = \frac{z}{1} - \frac{z^2}{2} + \frac{z^3}{3} - \cdots = \sum_{s=1}^{\infty} \frac{\varkappa_s}{s!}(it)^s. \tag{4.3.3}$$

Aus (4.3.2) und (4.3.3) erhalten wir formal

$$\varphi(t) = 1 + \sum_{s=1}^{\infty} \frac{m_s}{s!} (it)^s = \exp\left(\sum_{s=1}^{\infty} \frac{\varkappa_s}{s!} (it)^s\right)$$
$$= 1 + \sum_{s=1}^{\infty} \frac{\varkappa_s}{s!} (it)^s + \frac{1}{2!}\left[\sum_{s=1}^{\infty} \frac{\varkappa_s}{s!} (it)^s\right]^2 + \frac{1}{3!}\left[\sum_{s=1}^{\infty} \frac{\varkappa_s}{s!} (it)^s\right]^3 + \cdots. \tag{4.3.4}$$

Definition 4.3.1. Die in der Formel (4.3.3) auftretenden Koeffizienten $\varkappa_s$ nennen wir *Semiinvarianten.*

Um die Semiinvarianten durch die Momente bzw. die Momente durch die Semiinvarianten auszudrücken, vergleichen wir die Koeffizienten von $(it)^s$ in (4.3.4). Auf diese Weise erhalten wir

$$\begin{aligned}
\varkappa_1 &= m_1, \\
\varkappa_2 &= m_2 - m_1^2 = \sigma^2, \\
\varkappa_3 &= m_3 - 3m_1m_2 + 2m_1^3, \\
\varkappa_4 &= m_4 - 3m_2^2 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4, \\
&\cdots\cdots\cdots\cdots\cdots
\end{aligned} \tag{4.3.5}$$

und ebenso

$$\begin{aligned}
m_1 &= \varkappa_1, \\
m_2 &= \varkappa_2 + \varkappa_1^2, \\
m_3 &= \varkappa_3 + 3\varkappa_1\varkappa_2 + \varkappa_1^3, \\
m_4 &= \varkappa_4 + 3\varkappa_2^2 + 4\varkappa_1\varkappa_3 + 6\varkappa_1^2\varkappa_2 + \varkappa_1^4, \\
&\cdots\cdots\cdots\cdots\cdots
\end{aligned} \tag{4.3.6}$$

Die Semiinvarianten lassen sich auch durch die zentralen Momente ausdrücken:

$$\begin{aligned}
\varkappa_1 &= m_1, \\
\varkappa_2 &= \mu_2 = \sigma^2, \\
\varkappa_3 &= \mu_3, \\
\varkappa_4 &= \mu_4 - 3\mu_2^2, \\
&\cdots\cdots
\end{aligned} \tag{4.3.7}$$

Aus den Formeln (4.3.5) und (4.3.6) folgt: Existiert das Moment l-ter Ordnung, so existieren auch alle Semiinvarianten, deren Ordnung nicht größer als l ist.

Die Bezeichnung „Semiinvariante“ erklärt sich daraus, daß bei einer Translation der Zufallsvariablen X, d. h. bei einer Transformation $Y = X + b$, alle Semiinvarianten außer $\varkappa_1$ unverändert bleiben. In der Tat, wenn wir mit $\varphi(t)$ bzw. $\varphi_1(t)$ die charakteristischen Funktionen der Zufallsvariablen X bzw. Y bezeichnen, so folgt aus (4.2.14)

$$\log \varphi_1(t) = bit + \log \varphi(t). \tag{4.3.8}$$

Die Translation ändert nur den Koeffizienten von it, d. h. den Koeffizienten der ersten Potenz in der Entwicklung (4.3.4), also nur die Semiinvariante erster Ordnung.

Beispiel 4.3.1. Wir berechnen die Semiinvarianten der im Beispiel 4.2.1 besprochenen Poissonverteilung.

Die charakteristische Funktion der Poissonverteilung ist

$$\varphi(t) = e^{\lambda(e^{it}-1)}.$$

Daraus erhalten wir

$$\psi(t) = \log \varphi(t) = \lambda(e^{it} - 1) = \lambda\left(\sum_{k=0}^{\infty} \frac{(it)^k}{k!} - 1\right) = \lambda \sum_{k=1}^{\infty} \frac{(it)^k}{k!}. \tag{4.3.9}$$

Aus (4.3.3) ergibt sich

$$\varkappa_k = \lambda \quad (k = 1, 2, \ldots). \tag{4.3.10}$$

Benutzen wir die Beziehungen, die zwischen den Semiinvarianten und den Momenten bestehen, so können wir auf Grund der Formel (4.3.10) für die Poissonverteilung Momente beliebiger Ordnung berechnen.

4.4. Die charakteristische Funktion einer Summe unabhängiger Zufallsvariabler

Es seien X und Y unabhängige Zufallsvariable. Aus den Untersuchungen in 2.8 folgt, daß dann die Zufallsvariablen e^{itX} und e^{itY} ebenfalls unabhängig sind.

Wir wollen nun die charakteristische Funktion der Summe

$$Z = X + Y$$

von zwei Zufallsvariablen bestimmen.

Mit $\varphi(t)$, $\varphi_1(t)$, $\varphi_2(t)$ seien die charakteristischen Funktionen der Zufallsvariablen Z, X bzw. Y bezeichnet. Wir erhalten

$$\varphi(t) = E(e^{itZ}) = E(e^{it(X+Y)}) = E(e^{itX} e^{itY}). \tag{4.4.1}$$

Da die Zufallsvariablen e^{itX} und e^{itY} unabhängig sind, ergibt sich aus Satz 3.6.2

$$\varphi(t) = E(e^{itX}) E(e^{itY}) = \varphi_1(t) \varphi_2(t). \tag{4.4.2}$$

Dieses Ergebnis kann man auf die Summe endlich vieler unabhängiger Zufallsvariabler verallgemeinern.

Satz 4.4.1. *Die charakteristische Funktion der Summe endlich vieler unabhängiger Zufallsvariabler ist gleich dem Produkt der charakteristischen Funktionen dieser Zufallsvariablen.*

Wenn also Z die Summe von n unabhängigen Zufallsvariablen ist,

$$Z = X_1 + X_2 + \cdots + X_n,$$

und $\varphi(t)$ bzw. $\varphi_k(t)$ $(k = 1, 2, \ldots, n)$ die charakteristischen Funktionen der Zufallsvariablen Z bzw. X_k sind, erhalten wir

$$\varphi(t) = \varphi_1(t)\varphi_2(t)\cdots\varphi_n(t). \tag{4.4.3}$$

Beispiel 4.4.1. Wir betrachten zwei unabhängige Zufallsvariable X_1 und X_2 mit Poissonverteilungen. Dabei sei für $r = 0, 1, 2, \ldots$

$$P(X_1 = r) = \frac{\lambda_1^r}{r!} e^{-\lambda_1},$$

$$P(X_2 = r) = \frac{\lambda_2^r}{r!} e^{-\lambda_2}.$$

Wir suchen die charakteristische Funktion der Zufallsvariablen $Z = X_1 - X_2$ und die Semiinvarianten ihrer Verteilung.

Nach der Gleichung (4.2.6) sind die charakteristischen Funktionen $\varphi_1(t)$ und $\varphi_2(t)$ der Zufallsvariablen X_1 und X_2 gleich

$$\varphi_1(t) = e^{\lambda_1(e^{it}-1)},$$

$$\varphi_2(t) = e^{\lambda_2(e^{it}-1)}.$$

Die charakteristische Funktion der Zufallsvariablen $-X_2$ hat wegen der Beziehung (4.2.16) die Gestalt

$$\varphi_2(-t) = e^{\lambda_2(e^{-it}-1)}.$$

Berücksichtigen wir die Unabhängigkeit der Zufallsvariablen X_1 und $-X_2$ sowie die Gleichung (4.4.2), so erhalten wir als charakteristische Funktion der Zufallsvariablen Z den Ausdruck

$$\varphi(t) = e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{-it}-1)} = e^{\lambda_1 e^{it} + \lambda_2 e^{-it} - \lambda_1 - \lambda_2}.$$

Wenn wir die Exponenten e^{it} und e^{-it} in Potenzreihen entwickeln, erhalten wir

$$\varphi(t) = e^{(\lambda_1-\lambda_2)(it)+(\lambda_1+\lambda_2)\frac{(it)^2}{2!}+(\lambda_1-\lambda_2)\frac{(it)^3}{3!}+\cdots},$$

$$\psi(t) = \log\varphi(t) = (\lambda_1 - \lambda_2)\frac{(it)}{1!} + (\lambda_1 + \lambda_2)\frac{(it)^2}{2!} + (\lambda_1 - \lambda_2)\frac{(it)^3}{3!} + \cdots.$$

Aus den Gleichungen (4.3.3) folgt, daß alle ungeraden Semiinvarianten der Zufallsvariablen Z gleich $\lambda_1 - \lambda_2$, alle geraden gleich $\lambda_1 + \lambda_2$ sind. Aus den Formeln (4.3.5) erhalten wir für den Mittelwert und die Dispersion:

$$m_1 = \varkappa_1 = \lambda_1 - \lambda_2,$$

$$\sigma^2 = \varkappa_2 = \lambda_1 + \lambda_2.$$

Wir bemerken, daß die Umkehrung des Satzes 4.4.1 nicht gilt. Auch die charakteristische Funktion der Summe abhängiger Zufallsvariabler kann gleich dem Produkt ihrer charakteristischen Funktionen sein. Wir wollen das an einem Beispiel zeigen.

Beispiel 4.4.2. Die Verteilung der Zufallsvariablen (X, Y) sei durch die Dichte

$$f(x, y) = \begin{cases} \frac{1}{4}\,[1 + xy(x^2 - y^2)] & \text{für} \quad |x| \leqq 1 \quad \text{und} \quad |y| \leqq 1, \\ 0 & \text{für die übrigen Punkte} \end{cases}$$

definiert. Wir zeigen zunächst, daß X und Y abhängig sind.

Die Randdichten werden in den Gebieten $|x| \leqq 1$ und $|y| \leqq 1$ durch die Formeln

$$f_1(x) = \int_{-1}^{1} \frac{1}{4}\,[1 + xy(x^2 - y^2)]\,dy = \frac{1}{4}\left[y + \frac{x^3y^2}{2} - \frac{xy^4}{4}\right]_{-1}^{1} = \frac{1}{2},$$

$$f_2(y) = \int_{-1}^{1} \frac{1}{4}\,[1 + xy(x^2 - y^2)]\,dx = \frac{1}{4}\left[x + \frac{x^4y}{4} - \frac{x^2y^3}{2}\right]_{-1}^{1} = \frac{1}{2}$$

gegeben. Daraus folgt $f_1(x)f_2(y) = \frac{1}{4} \neq f(x,y)$; die Variablen X und Y sind also nicht unabhängig.

Nun suchen wir die Dichte der Summe $Z = X + Y$. Aus (2.9.8) ergibt sich

$$f_3(z) = \int_{-\infty}^{\infty} f(x, z - x)\,dx.$$

Um die von z abhängigen Integrationsgrenzen dieses Integrals zu finden, beachten wir, daß bei der Einführung der Zufallsvariablen x,z statt x,y das quadratische Gebiet $|x| \leqq 1$, $|y| \leqq 1$ in ein durch die Ungleichungen

$$|x| \leqq 1, \quad x - 1 \leqq z \leqq x + 1 \tag{4.4.4}$$

bestimmtes Gebiet übergeht. Diese Transformation wird durch Abb. 4.4.1 und 4.4.2 illustriert. In Abb. 4.4.1 ist das Gebiet $|x| \leqq 1$, $|y| \leqq 1$ der x,y-Ebene dargestellt, in Abb. 4.4.2 das Gebiet der x,z-Ebene, das aus dem vorhergehenden durch die Transformation $x = x, z = x + y$ hervorgeht. Wir schreiben die Ungleichungen (4.4.4) in der Gestalt

$$|x| \leqq 1, \quad z - 1 \leqq x \leqq z + 1.$$

Dann gelten für $z \leqq 0$ die Ungleichungen

$$z - 1 \leqq -1, \quad z + 1 \leqq 1$$

und für $z > 0$ die Ungleichungen

$$z - 1 > -1, \quad z + 1 > 1.$$

Wir haben also für $z \leqq 0$ in den Grenzen von -1 bis $z + 1$ und für $z > 0$ in den Grenzen von $z - 1$ bis 1 zu integrieren. Nach einfachen Umformungen erhalten wir

$$f_3(z) = \begin{cases} \displaystyle\int_{-1}^{z+1} \frac{1}{4}\,(1 + 3z^2x^2 - 2zx^3 - z^3x)\,dx = \frac{1}{4}\,(2 + z) & \text{für} \quad -2 \leqq z \leqq 0, \\ \displaystyle\int_{z-1}^{1} \frac{1}{4}\,(1 + 3z^2x^2 - 2zx^3 - z^3x)\,dx = \frac{1}{4}\,(2 - z) & \text{für} \quad 0 < z \leqq 2, \\ 0 & \text{für } |z| > 2. \end{cases}$$

Abb. 4.4.1

Abb. 4.4.2

Eine Verteilung, wie sie die Zufallsvariable Z besitzt, nennen wir *Dreiecksverteilung*. In Abb. 4.4.3 ist die Funktion $f_3(z)$ dargestellt.

Wir bestimmen nun die charakteristischen Funktionen $\varphi_1(t)$, $\varphi_2(t)$ und $\varphi_3(t)$ der Zufallsvariablen X, Y und $Z = X + Y$. Dann erhalten wir

$$\varphi_1(t) = \frac{1}{2}\int_{-1}^{1} e^{itx}\,dx = \frac{1}{2}\left[\frac{e^{itx}}{it}\right]_{-1}^{1} = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}$$

und ebenso

$$\varphi_2(t) = \frac{\sin t}{t}.$$

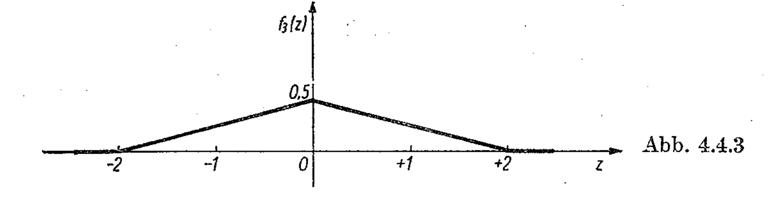


Abb. 4.4.3

Berücksichtigen wir, daß die Zufallsvariable Z nur Werte aus dem Intervall $[-2, 2]$ annimmt, so finden wir

$$\varphi_3(t) = \frac{1}{4}\int_{-2}^{0} (2+z)e^{itz}\,dz + \frac{1}{4}\int_{0}^{2} (2-z)e^{itz}\,dz = \frac{1}{4}\left[\frac{2-e^{2it}-e^{-2it}}{t^2}\right]$$

$$= \frac{1}{2t^2}\left(1 - \frac{e^{2it}+e^{-2it}}{2}\right) = \frac{1}{2t^2}(1-\cos 2t) = \left(\frac{\sin t}{t}\right)^2.$$

Daraus folgt, daß die Gleichung $\varphi_3(t) = \varphi_1(t)\varphi_2(t)$ gilt, obwohl die Zufallsvariablen X und Y abhängig sind.

4.5. Die Bestimmung der Verteilungsfunktion durch die charakteristische Funktion

A. Durch die Formel (4.1.1) wird zu einer gegebenen Verteilung eindeutig die charakteristische Funktion definiert. Man kann zeigen, daß auch das Umgekehrte der Fall ist: Durch die charakteristische Funktion ist die Verteilungsfunktion eindeutig bestimmt. Dazu beweisen wir den Satz von Lévy [3].

Satz 4.5.1. *Es seien $F(x)$ bzw. $\varphi(t)$ die Verteilungsfunktion bzw. die charakteristische Funktion der Zufallsvariablen X. Sind $a+h$ und $a-h$ $(h > 0)$ beliebige Stetigkeitsstellen der Verteilungsfunktion $F(x)$, so ist*

$$F(a+h) - F(a-h) = \lim_{T\to\infty} \frac{1}{\pi}\int_{-T}^{T} \frac{\sin ht}{t}\, e^{-ita}\varphi(t)\,dt. \tag{4.5.1}$$

Bevor wir zum Beweis kommen, wollen wir die Bedeutung dieses Satzes erläutern. Da die Größen a und h beliebig sind, erhält man aus der Formel (4.5.1) für beliebige Stetigkeitsstellen x_1 und x_2 die Differenz $F(x_2) - F(x_1)$. Wegen der Beziehung

$$F(x_2) - F(x_1) = P(x_1 \leqq X < x_2)$$

erlaubt uns also der Satz 4.5.1, die Wahrscheinlichkeit dafür zu berechnen, daß der Wert von X einem beliebigen Intervall angehört, dessen Endpunkte Stetigkeitsstellen der Verteilungsfunktion $F(x)$ sind, wenn wir die charakteristische Funktion $\varphi(t)$ kennen. Es sei $x = x_2$ eine Stetigkeitsstelle. Wir betrachten auf der Menge der Stetigkeitsstellen den Grenzübergang $x_1 \to -\infty$. Die Folge der Differenzen $F(x) - F(x_1)$ kann dann aus der charakteristischen Funktion berechnet werden. Da sie gegen $F(x)$ konvergiert, kann die Verteilungsfunktion $F(x)$ in einer beliebigen Stetigkeitsstelle bestimmt werden und ist damit überall bestimmt.

Wir kommen nun zum Beweis des Satzes 4.5.1.

Beweis. Wir führen den Beweis nur für eine stetige Zufallsvariable mit der Dichtefunktion $f(x)$. Es sei

$$J = \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t} e^{-ita} \varphi(t)\, dt. \tag{4.5.2}$$

Wenn wir die Definition der charakteristischen Funktion beachten, erhalten wir

$$\begin{aligned} J &= \frac{1}{\pi} \int_{-T}^{T} \int_{-\infty}^{\infty} \frac{\sin ht}{t} e^{-ita} e^{itx} f(x)\, dx\, dt \\ &= \frac{1}{\pi} \int_{-T}^{T} \int_{-\infty}^{\infty} \frac{\sin ht}{t} e^{it(x-a)} f(x)\, dx\, dt. \end{aligned}$$

In diesem Integral können wir die Reihenfolge der Integration ändern, weil die Integrationsgrenzen bezüglich t endlich sind und das Integral bezüglich x absolut konvergent ist. Es ist nämlich

$$\int_{-\infty}^{\infty} \left| \frac{\sin ht}{t} e^{it(x-a)} \right| f(x)\, dx = \int_{-\infty}^{\infty} \left| \frac{\sin ht}{t} \right| f(x)\, dx \leqq h \int_{-\infty}^{\infty} f(x)\, dx = h.$$

Wir erhalten also

$$\begin{aligned} J &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{\sin ht}{t} e^{it(x-a)} f(x)\, dt\, dx \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{\sin ht}{t} \{\cos [(x-a)t] + i \sin [(x-a)t]\} f(x)\, dt\, dx \\ &= \frac{2}{\pi} \int_{-\infty}^{\infty} \int_{0}^{T} \frac{\sin ht}{t} \cos [(x-a)t] f(x)\, dt\, dx. \end{aligned}$$

Wenn wir die Formel

$$\sin A \cos B = \frac{\sin (A+B) + \sin (A-B)}{2}$$

anwenden und $A = ht$, $B = (x - a)t$ setzen, erhalten wir

$$J = \int_{-\infty}^{\infty} \left[\frac{1}{\pi} \int_0^T \frac{\sin(x - a + h)t}{t}\, dt - \frac{1}{\pi} \int_0^T \frac{\sin(x - a - h)t}{t}\, dt \right] f(x)\, dx$$

$$= \int_{-\infty}^{\infty} g(x, T) f(x)\, dx, \qquad (4.5.2')$$

wobei $g(x, T)$ der Ausdruck in der eckigen Klammer ist. Aus der Analysis ist bekannt, daß das Integral $\int_0^T \frac{\sin x}{x}\, dx$ für alle $T > 0$ beschränkt ist und für $T \to \infty$ gegen $\frac{\pi}{2}$ konvergiert. Daraus folgt, daß auch der Ausdruck $|g(x,T)|$ beschränkt ist und daß

$$\lim_{T\to\infty} \frac{1}{\pi} \int_0^T \frac{\sin \alpha t}{t}\, dt = \begin{cases} \frac{1}{2} & \text{für} \quad \alpha > 0, \\ -\frac{1}{2} & \text{für} \quad \alpha < 0. \end{cases}$$

Dabei ist die Konvergenz gleichmäßig in α für

$$|\alpha| = |x - a \pm h| > \delta > 0.$$

Wenn wir dieses Ergebnis beachten, erhalten wir

$$\lim_{T\to\infty} g(x,T) = \begin{cases} 0 & \text{für} \quad x < a - h, \\ \frac{1}{2} & \text{für} \quad x = a - h, \\ 1 & \text{für} \quad a - h < x < a + h, \\ \frac{1}{2} & \text{für} \quad x = a + h, \\ 0 & \text{für} \quad x > a + h. \end{cases} \qquad (4.5.3)$$

Daraus folgt: Bei der Berechnung vom $\lim_{T\to\infty} J$ kann man unter dem Integral auf der rechten Seite der Formel (4.5.2′) zur Grenze übergehen.

Wenn wir noch berücksichtigen, daß die Verteilungsfunktion $F(x)$ in den Punkten $x = a - h$ und $x = a + h$ stetig ist, erhalten wir

$$\lim_{T\to\infty} J = \int_{-\infty}^{\infty} \lim_{T\to\infty} g(x, T) f(x)\, dx = \int_{a-h}^{a+h} f(x)\, dx = F(a + h) - F(a - h). \qquad (4.5.4)$$

Aus (4.5.2) und (4.5.4) ergibt sich die Formel (4.5.1). Der Satz ist damit für stetige Variable bewiesen. Für diskrete Variable ist der Beweis analog; die Integrale sind nur durch Reihen zu ersetzen.

B. Ist die charakteristische Funktion $\varphi(t)$ im Intervall $(-\infty, \infty)$ absolut integrierbar, dann kann auch die ihr entsprechende Dichtefunktion $f(x)$ durch die Funktion $\varphi(t)$ bestimmt werden.[1] Aus der absoluten Integrierbarkeit der Funktion $\varphi(t)$ folgt nämlich, daß das uneigentliche Integral (4.5.1) existiert. Wir dividieren beide Seiten der Gleichung (4.5.1) durch $2h$ und erhalten

$$\frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin ht}{ht} e^{-itx} \varphi(t)\, dt, \tag{4.5.5}$$

wobei $x + h$ und $x - h$ Stetigkeitsstellen von $F(x)$ sind. Strebt h gegen Null, dann strebt der Ausdruck unter dem Integral gegen $e^{-itx}\varphi(t)$. Außerdem ist der absolute Betrag des Ausdrucks unter dem Integral nicht größer als $|\varphi(t)|$; es sollte aber $\varphi(t)$ nach Voraussetzung absolut integrierbar sein. Daraus folgt, daß man auf der rechten Seite der Gleichung (4.5.5) unter dem Integral zur Grenze für $h \to 0$ übergehen kann. Wir erhalten dann

$$\lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t)\, dt.$$

Da die rechte Seite dieser Gleichung eine stetige Funktion von x ist, erhalten wir daraus

$$F'(x) = f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t)\, dt. \tag{4.5.6}$$

Aus der absoluten und gleichmäßigen Konvergenz dieses Integrals ergibt sich, daß die Dichte $F'(x)$ eine stetige Funktion ist. Also erlaubt die Formel (4.5.6), die Dichte $f(x)$ mit Hilfe der charakteristischen Funktion $\varphi(t)$ zu bestimmen, wenn $\varphi(t)$ absolut integrierbar ist. Wie man sieht, ist (4.5.6) die Umkehrung der Formel (4.1.4).

Beispiel 4.5.1. Die charakteristische Funktion der Variablen X sei durch

$$\varphi(t) = e^{-\frac{t^2}{2}}$$

gegeben. Gesucht ist die Dichte dieser Variablen.

[1]) Wie VAN DER VAART [2] kürzlich gezeigt hat, gilt dies auch bei schwächeren Voraussetzungen über die Funktion $\varphi(t)$.

Aus Gleichung (4.5.6) ergibt sich

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} e^{-\frac{t^2}{2}} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{(t+ix)^2}{2}} e^{\frac{(ix)^2}{2}} dt$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(t+ix)^2}{2}} dt = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Dies ist die *Dichte der Gaußschen Verteilung.*

Der Leser vergleiche dieses Ergebnis mit dem Ergebnis (4.2.10) im Beispiel 4.2.2, wo wir die umgekehrte Fragestellung betrachteten; als Lösung fanden wir die charakteristische Funktion einer Gaußschen Verteilung.

Ist die Zufallsvariable X diskret und nimmt sie nur ganzzahlige Werte an, dann kann man leicht ihre Wahrscheinlichkeitsfunktion durch die charakteristische Funktion bestimmen.

Für jedes ganzzahlige k sei

$$p_k = P(X = k),$$

wobei selbstverständlich nicht alle Wahrscheinlichkeiten positiv sein müssen. Dann ist

$$\varphi(t) = \sum_{k=-\infty}^{\infty} p_k e^{ikt}.$$

Ist k' eine bestimmte ganze Zahl, so erhalten wir

$$e^{-itk'} \varphi(t) = \sum_{\substack{k=-\infty \\ k \neq k'}}^{\infty} e^{-it(k'-k)} p_k + p_{k'}.$$

Wenn wir beide Seiten dieser Gleichung in den Grenzen von $-\pi$ bis π integrieren und beachten, daß für jedes $k \neq k'$

$$\int_{-\pi}^{\pi} e^{-it(k'-k)} dt = 0$$

ist, so erhalten wir nach Weglassen des Strichs bei k'

$$p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \varphi(t)\, dt. \tag{4.5.7}$$

Der Leser beachte die Analogie der Formeln (4.5.6) und (4.5.7).

C. GNEDENKO [1] hat bewiesen, daß durch die Werte, die die charakteristischen Funktionen auf einem endlichen Intervall annimmt, *die Verteilungsfunktion noch nicht eindeutig bestimmt ist.* Dies wird das folgende Beispiel zeigen.

Beispiel 4.5.2. Wir wollen die Dichte $f(x)$ der Zufallsvariablen X bestimmen, deren charakteristische Funktion folgendermaßen definiert ist:

$$\varphi_1(t) = \begin{cases} 1 - |t| & \text{für } |t| \leqq 1, \\ 0 & \text{für } |t| > 1. \end{cases} \tag{4.5.8}$$

Offensichtlich ist die Funktion $\varphi_1(t)$ im Intervall $-\infty < t < \infty$ absolut integrierbar. Aus Formel (4.5.6) ergibt sich

$$f(x) = \frac{1}{2\pi} \int_{\infty}^{\infty} e^{-itx} \varphi_1(t)\, dt = \frac{1}{2\pi} \int_{-1}^{0} (1 + t) e^{-itx} dt + \frac{1}{2\pi} \int_{0}^{1} (1 - t) e^{-itx} dt,$$

$$\int_{-1}^{0} (1 + t) e^{-itx} dt = \left[\frac{e^{-itx}}{-ix} (1 + t) \right]_{-1}^{0} - \frac{1}{-ix} \int_{-1}^{0} e^{-itx} dt$$

$$= -\frac{1}{ix} + \frac{1}{ix} \left[\frac{e^{-itx}}{-ix} \right]_{-1}^{0} = -\frac{1}{ix} - \frac{1}{(ix)^2} (1 - e^{ix}),$$

$$\int_{0}^{1} (1 - t) e^{-itx} dt = \left[\frac{e^{-itx}}{-ix} (1 - t) \right]_{0}^{1} + \frac{1}{-ix} \int_{0}^{1} e^{-itx} dt$$

$$= \frac{1}{ix} - \frac{1}{ix} \left[\frac{e^{-itx}}{-ix} \right]_{0}^{1} = \frac{1}{ix} + \frac{1}{(ix)^2} (e^{-ix} - 1).$$

Daraus erhält man

$$f(x) = \frac{1}{2\pi x^2} (2 - e^{ix} - e^{-ix}) = \frac{1}{\pi x^2} \left(1 - \frac{e^{ix} + e^{-ix}}{2} \right) = \frac{1 - \cos x}{\pi x^2}. \tag{4.5.9}$$

Nun wollen wir die diskrete Zufallsvariable Y mit der Wahrscheinlichkeitsfunktion

$$P(Y = 0) = \frac{1}{2},$$

$$P[Y = (2k - 1)\pi] = \frac{2}{(2k - 1)^2 \pi^2} \quad (k = 0, \pm 1, \pm 2, \ldots) \tag{4.5.10}$$

betrachten. Als charakteristische Funktion $\varphi_2(t)$ dieser Verteilung finden wir

$$\varphi_2(t) = \frac{1}{2} + \sum_{k=-\infty}^{\infty} \frac{2}{(2k - 1)^2 \pi^2} e^{it(2k-1)\pi}$$

$$= \frac{1}{2} + \frac{2}{\pi^2} \sum_{k=-\infty}^{\infty} \frac{\cos(2k - 1) t\pi + i \sin(2k - 1) t\pi}{(2k - 1)^2}$$

$$= \frac{1}{2} + \frac{4}{\pi^2} \sum_{k=1}^{\infty} \frac{\cos(2k - 1) t\pi}{(2k - 1)^2}. \tag{4.5.11}$$

Wie wir zeigen werden, gilt für $|t| \leqq 1$ die Beziehung $\varphi_1(t) = \varphi_2(t)$.

Die Funktion $\psi(t) = |t|$ entwickeln wir im Intervall $|t| \leqq 1$ in eine Fourierreihe:

$$\psi(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\pi t.$$

Die Koeffizienten dieser Entwicklung berechnen wir aus den Formeln

$$\frac{a_0}{2} = \int_0^1 t\,dt = \frac{1}{2},$$

$$a_n = 2\int_0^1 t \cos n\pi t\,dt = \left[\frac{2t \sin n\pi t}{n\pi}\right]_0^1 - \frac{2}{n\pi}\int_0^1 \sin n\pi t\,dt$$

$$= -\frac{2}{n\pi}\left[\frac{-\cos n\pi t}{n\pi}\right]_0^1 = 2\,\frac{\cos n\pi - 1}{\pi^2 n^2}.$$

Für gerade n ist $a_n = 0$, und für ungerade, d. h. für $n = 2k - 1$, ist

$$a_{2k-1} = -\frac{4}{(2k-1)^2\pi^2}.$$

Endgültig erhalten wir

$$\psi(t) = |t| = \frac{1}{2} - \frac{4}{\pi^2}\sum_{k=1}^{\infty}\frac{\cos(2k-1)\pi t}{(2k-1)^2} \tag{4.5.12}$$

und aus den Formeln (4.5.11) und (4.5.12)

$$\varphi_2(t) = 1 - |t| = \varphi_1(t),$$

obwohl $\varphi_1(t)$ und $\varphi_2(t)$ charakteristische Funktionen verschiedener Verteilungen sind. Für $|t| > 1$ sind die charakteristischen Funktionen $\varphi_1(t)$ und $\varphi_2(t)$ nicht gleich, denn nach Definition ist dort $\varphi_1(t) = 0$, während die Funktion $\varphi_2(t)$ nicht identisch gleich Null ist; die Werte, die diese Funktion im Intervall $|t| \leqq 1$ annimmt, wiederholen sich nämlich auf der ganzen t-Achse periodisch.

4.6. Die charakteristische Funktion einer mehrdimensionalen Zufallsvariablen

A. Der Begriff der charakteristischen Funktion einer eindimensionalen Zufallsvariablen kann auf Zufallsvariable mit endlich vielen Dimensionen ausgedehnt werden. Wir beschränken uns hier auf zweidimensionale Zufallsvariable.

Es sei (X, Y) eine zweidimensionale Zufallsvariable, t und u seien beliebige reelle Zahlen. Die charakteristische Funktion der Zufallsvariablen (X, Y) wird durch

$$\varphi(t, u) = E(e^{i(tX+uY)}) \tag{4.6.1}$$

definiert.

Beispiel 4.6.1. Die zweidimensionale Zufallsvariable (X, Y) soll die vier Zahlenpaare $(1, 1)$ $(1, -1)$ $(-1, 1)$ $(-1, -1)$ mit den Wahrscheinlichkeiten

$$P(X = 1, Y = 1) = \frac{1}{3}, \qquad P(X = 1, Y = -1) = \frac{1}{3},$$

$$P(X = -1, Y = 1) = \frac{1}{6}, \qquad P(X = -1, Y = -1) = \frac{1}{6}$$

annehmen können.

Wie sich der Leser leicht überzeugt, sind die Zufallsvariablen X und Y unabhängig. Wir berechnen die charakteristische Funktion der Zufallsvariablen (X, Y) nach (4.6.1) und erhalten

$$\begin{aligned}\varphi(t,u) &= E(e^{i(tX+uY)}) = \frac{1}{3}\, e^{i(t+u)} + \frac{1}{3}\, e^{i(t-u)} + \frac{1}{6}\, e^{i(-t+u)} + \frac{1}{6}\, e^{i(-t-u)} \\ &= \frac{1}{3}\, e^{it}(e^{iu} + e^{-iu}) + \frac{1}{6}\, e^{-it}(e^{iu} + e^{-iu}) \\ &= \frac{1}{6}\,(e^{iu} + e^{-iu})\,(2e^{it} + e^{-it}) = \frac{1}{3}\cos u\,(3\cos t + i\sin t). \qquad (4.6.2)\end{aligned}$$

Jetzt wollen wir einige Eigenschaften der mehrdimensionalen charakteristischen Funktionen untersuchen:

$$\varphi(0, 0) = E(e^{i(0X+0Y)}) = 1, \qquad (4.6.3)$$

$$|\varphi(t, u)| = |E(e^{i(tX+uY)})| \leqq E(|e^{i(tX+uY)}|) = 1,$$

also

$$|\varphi(t, u)| \leqq 1, \qquad (4.6.4)$$

$$\varphi(-t, -u) = E(e^{-i(tX+uY)}) = \overline{\varphi(t, u)}. \qquad (4.6.5)$$

Ebenso wie für eindimensionale Zufallsvariable kann man zeigen: Existieren alle Momente k-ter Ordnung der Zufallsvariablen, so existieren auch die Ableitungen

$$\frac{\partial^k \varphi(t, u)}{\partial t^{k-l}\, \partial u^l} \quad \text{für} \quad l = 0, 1, 2, \ldots, k, \qquad (4.6.6)$$

und man kann sie aus der Formel

$$\frac{\partial^k \varphi(t, u)}{\partial t^{k-l}\, \partial u^l} = i^k E(X^{k-l}\, Y^l\, e^{i(tX+uY)}) \qquad (4.6.7)$$

erhalten. Aus dieser Gleichung erkennt man, daß man die Momente $m_{k-l,l}$ aus der Formel

$$m_{k-l,l} = E(X^{k-l}\, Y^l) = \frac{1}{i^k} \left[\frac{\partial^k \varphi(t, u)}{\partial t^{k-l}\, \partial u^l}\right]_{\substack{t=0\\u=0}} \qquad (4.6.8)$$

berechnen kann. Damit erhalten wir für die Momente erster und zweiter Ordnung die Ausdrücke

$$m_{10} = \frac{1}{i}\left[\frac{\partial \varphi(t, u)}{\partial t}\right]_{\substack{t=0\\u=0}}, \qquad m_{01} = \frac{1}{i}\left[\frac{\partial \varphi(t, u)}{\partial u}\right]_{\substack{t=0\\u=0}},$$

$$m_{20} = \frac{1}{i^2}\left[\frac{\partial^2 \varphi(t, u)}{\partial t^2}\right]_{\substack{t=0\\u=0}}, \qquad m_{11} = \frac{1}{i^2}\left[\frac{\partial^2 \varphi(t, u)}{\partial t\, \partial u}\right]_{\substack{t=0\\u=0}},$$

$$m_{02} = \frac{1}{i^2}\left[\frac{\partial^2 \varphi(t, u)}{\partial u^2}\right]_{\substack{t=0\\u=0}}. \tag{4.6.8'}$$

Die charakteristischen Funktionen der Randverteilungen der Zufallsvariablen X bzw. Y erhalten wir aus den Gleichungen (4.6.1), wenn wir $u = 0$ bzw. $t = 0$ setzen. Es ist

$$\varphi(t, 0) = E(e^{itX}) = \varphi_1(t). \tag{4.6.9}$$

Das ist die charakteristische Funktion der Zufallsvariablen X. Analog ist

$$\varphi(0, u) = E(e^{itY}) = \varphi_2(u) \tag{4.6.10}$$

die charakteristische Funktion der Zufallsvariablen Y.

B. Wir wollen nun ohne Beweis die Verallgemeinerung des Satzes 4.5.1 für zweidimensionale Zufallsvektoren angeben. Der Beweis ist dem des Satzes für eine eindimensionale Zufallsvariable ähnlich.

Satz 4.6.1. *Es sei $\varphi(t, u)$ die charakteristische Funktion der Zufallsvariablen (X, Y). Ist das Rechteck $(a - h \leqq X < a + h,\ b - g \leqq Y < b + g)$ ein Stetigkeitsrechteck* (vgl. Definition 2.5.6), *so gilt*

$$P(a - h \leqq X < a + h, b - g \leqq Y < b + g)$$

$$= \lim_{T\to\infty} \frac{1}{\pi^2} \int_{-T}^{T} \int_{-T}^{T} \frac{\sin ht}{t} \frac{\sin gu}{u} \exp\big(-i(at + bu)\big)\varphi(t, u)\, dt\, du. \tag{4.6.11}$$

Die Formel (4.6.11) gestattet uns somit, falls $\varphi(t, u)$ bekannt ist, die Wahrscheinlichkeit

$$P(x_1 \leqq X < x_2, y_1 \leqq Y < y_2) \tag{4.6.12}$$

für ein beliebiges Stetigkeitsrechteck zu berechnen. Die Wahrscheinlichkeiten (4.6.12) für Stetigkeitsrechtecke bestimmen aber vollständig die Wahrscheinlichkeitsverteilung in der x, y-Ebene.

Satz 4.6.2. *Es seien $F(x, y)$, $F_1(x)$, $F_2(y)$, $\varphi(t, u)$, $\varphi_1(t)$ und $\varphi_2(u)$ die Verteilungsfunktionen und die charakteristischen Funktionen der Zufallsvariablen (X, Y), X und Y. Die Zufallsvariablen X und Y sind dann und nur dann unabhängig, wenn die Gleichung*

$$\varphi(t, u) = \varphi_1(t)\,\varphi_2(u) \tag{4.6.13}$$

für beliebige reelle t und u erfüllt ist.

Beweis. Es seien X und Y unabhängig. Nach Satz 3.2.2 gilt für beliebige reelle t und u

$$\varphi(t, u) = E(e^{i(tX+uY)}) = E(e^{itX})E(e^{iuY}) = \varphi_1(t)\varphi_2(u).$$

Nun sei die Gleichung (4.6.13) erfüllt. Dann erhalten wir aus (4.6.13), (4.6.11) und (4.5.1) die Gleichung

$$P(x_1 \leqq X < x_2, y_1 \leqq Y < y_2) = P(x_1 \leqq X < x_2)P(y_1 \leqq Y < y_2), \tag{4.6.14}$$

die für beliebige Stetigkeitsrechtecke gilt. Aus (4.6.14) folgt für beliebige x und y

$$F(x, y) = F_1(x)F_2(y).$$

Damit ist der Satz bewiesen.

Der nachstehende Satz von Cramér-Wold [1] ist für die Theorie der Verteilung von Zufallsvektoren von Bedeutung.

Satz 4.6.3. *Die Verteilungsfunktion $F(x, y)$ einer zweidimensionalen Zufallsvariablen (X, Y) ist eindeutig bestimmt durch die Klasse aller eindimensionalen Verteilungsfunktionen der eindimensionalen Zufallsvariablen $tX + uY$, wobei t und u die Menge aller reellen Zahlen durchlaufen.*

Beweis. Für alle reellen t und u seien die charakteristischen Funktionen $\varphi_z(v)$ von $Z = tX + uY$ gegeben:

$$\varphi_z(v) = E\big[\exp\big(iv(tX + uY)\big)\big] = E\big[\exp\big(i(vtX + vuY)\big)\big]. \tag{4.6.15}$$

Setzen wir $v = 1$, so erhalten wir auf der rechten Seite von (4.6.15) den Ausdruck

$$\varphi(t, u) = E[\exp(i(tX + uY))],$$

der die charakteristische Funktion der Verteilungsfunktion $F(x, y)$ darstellt. Nach Satz 4.6.1 definiert die Funktion $\varphi(t, u)$ eindeutig die Funktion $F(x, y)$. Damit ist der Satz bewiesen.

Wir schreiben

$$P(tX + uY < z) = P(X\cos\alpha + Y\sin\alpha < w)$$

mit

$$\cos\alpha = \frac{t}{\sqrt{t^2+u^2}}, \quad \sin\alpha = \frac{u}{\sqrt{t^2+u^2}}, \quad w = \frac{z}{\sqrt{t^2+u^2}} \quad (0 \leqq \alpha \leqq 2\pi).$$

Der Satz von CRAMÉR-WOLD läßt sich jetzt wie folgt formulieren: *Die Verteilungsfunktion $F(x, y)$ ist eindeutig bestimmt durch die Verteilungsfunktion der Projektionen von (X, Y) auf alle durch den Koordinatenursprung hindurchgehenden Geraden.*

Wie RÉNYI [5] gezeigt hat, gilt die folgende Behauptung: Genügt (X, Y) mit der Wahrscheinlichkeit 1 der Ungleichung

$$X^2 + Y^2 \leqq R^2 < \infty,$$

so ist die Verteilungsfunktion $F(x, y)$ durch die Verteilungsfunktionen der Zufallsvariablen $X\cos\alpha + Y\sin\alpha$ eindeutig bestimmt, wobei α eine abzählbar unendliche Menge verschiedener Werte aus dem Intervall $[0, 2\pi]$ durchläuft. Dem Problem, eine zweidimensionale Verteilung durch die Verteilungen der Projektionen auf gewisse abzählbare Mengen von Geraden zu bestimmen, sind auch die Arbeiten von GILBERT [1] und HEPPES [1] gewidmet.

Die Sätze 4.6.1 bis 4.6.3 lassen sich offenbar leicht auf Zufallsvektoren mit mehr als zwei Dimensionen verallgemeinern. Eine ausführliche Theorie der charakteristischen Funktionen findet der Leser im Buch von LUKACS [5].

4.7. Erzeugende Funktionen

Bei der Untersuchung von Zufallsvariablen, die ausschließlich ganzzahlige Werte $k = 0, 1, 2, \ldots$ annehmen, ist es vorteilhaft, an Stelle der charakteristischen Funktionen erzeugende Funktionen zu benutzen. X sei eine Zufallsvariable, und es sei

$$p_k = P(X = k) \quad (k = 0, 1, 2, \ldots)$$

mit $\sum_k p_k = 1$.

Definition 4.7.1. Die durch

$$\psi(s) = \sum_k p_k s^k \quad (-1 \leqq s \leqq 1) \tag{4.7.1}$$

definierte Funktion $\psi(s)$ nennen wir die *erzeugende Funktion* der ganzzahligen Zufallsvariablen X.

Wir bemerken, daß $\psi(1) = \sum_k p_k = 1$ ist; also ist die Reihe auf der rechten Seite von (4.7.1) im Intervall $|s| \leqq 1$ absolut und gleichmäßig konvergent. Die erzeugende Funktion ist dann stetig. Sie bestimmt eindeutig die Wahrscheinlichkeitsverteilung der Zufallsvariablen X, d. h. die Zahlen p_k, da $\psi(s)$ sich nur auf eine Weise als Potenzreihe der Form (4.7.1) darstellen läßt.

Beispiel 4.7.1. Die Zufallsvariable X habe eine Binomialverteilung:

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, \ldots, n).$$

Dann ist

$$\psi(s) = \sum_{k=0}^{n} \binom{n}{k} (ps)^k (1-p)^{n-k} = (ps+q)^n. \tag{4.7.2}$$

Beispiel 4.7.2. Die Zufallsvariable X habe eine Poissonverteilung:

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!} \qquad (k = 0, 1, \ldots).$$

Dann ist

$$\psi(s) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}. \tag{4.7.3}$$

Existiert das l-te Moment der Zufallsvariablen X, dann läßt es sich mit Hilfe der Ableitung der erzeugenden Funktion an der Stelle $s = 1$ bestimmen. Wir erhalten nämlich

$$\psi'(s) = \sum_k k p_k s^{k-1},$$

$$\psi''(s) = \sum_k k(k-1) p_k s^{k-2},$$

also

$$\psi'(1) = \sum_k k p_k = E(X),$$

$$\psi''(1) = \sum_k k(k-1) p_k = E(X^2) - E(X).$$

Daraus ergibt sich

$$E(X^2) = \psi''(1) + \psi'(1).$$

Für erzeugende Funktionen gilt ein zu dem Satz 4.4.1 analoger Satz. Sein Beweis ist elementar.

4.8. Aufgaben und Ergänzungen

1. Man bestimme die charakteristische Funktion der Zufallsvariablen, deren Dichtefunktionen die folgende Form haben:

$$\text{a) } f(x) = \begin{cases} 0 & \text{für } |x| \geqq a > 0, \\ \dfrac{a-1}{a^2} & \text{für } |x| < a; \end{cases}$$

$$\text{b) } f(x) = \frac{2 \sin^2 \frac{1}{2} a x}{\pi a x^2}.$$

Hinweis. Vergleiche hierzu die Herleitung der Formel (5.10.10).

2. Man zeige, daß die Funktion

$$\varphi(t) = \exp(-|t|^r),$$

wobei $r > 2$ ist, keine charakteristische Funktion darstellt.

3. Eine beschränkte und stetige Funktion $\varphi(t)$ ist charakteristische Funktion einer gewissen Verteilungsfunktion dann und nur dann, wenn $\varphi(0) = 1$ gilt und die Funktion

$$\psi(x, A) = \int_0^A \int_0^A \varphi(t-u) e^{ix(t-u)}\, dt\, du$$

für alle reellen x und alle $A > 0$ reell und nichtnegativ ist (Cramér [5]).

4. Die charakteristische Funktion einer Zufallsvariablen X sei $\varphi(t)$. Man beweise:

a) Ist X stetig, so ist $\lim\limits_{|t|\to\infty} \varphi(t) = 0$.

b) Ist X diskret, so ist $\limsup\limits_{|t|\to\infty} |\varphi(t)| = 1$.

5. Man beweise die folgenden Behauptungen:

a) Ist $\varphi(t)$ charakteristische Funktion einer gewissen Zufallsvariablen, so gilt das gleiche auch für $|\varphi(t)|^2$.

b) Ist $\varphi(t)$ charakteristische Funktion einer Zufallsvariablen mit stetiger Verteilungsfunktion, so hat $|\varphi(t)|^2$ die gleiche Eigenschaft.

6. Man zeige, daß die charakteristische Funktion einer Zufallsvariablen X dann und nur dann reell ist, wenn X eine symmetrische Verteilung bezüglich 0 besitzt.

7. a) Man zeige, daß für jede charakteristische Funktion $\varphi(t)$ die Beziehung

$$1 - \operatorname{Re}\varphi(2t) \leqq 4\bigl(1 - \operatorname{Re}\varphi(t)\bigr)$$

gilt.

b) Hieraus ist die Beziehung

$$1 - |\varphi(2t)|^2 \leqq 4\bigl(1 - |\varphi(t)|^2\bigr)$$

herzuleiten.

8. Man zeige, daß die Momente der Ordnungen $1, 2, \ldots, 2l$ einer Zufallsvariablen X existieren, wenn die charakteristische Funktion $\varphi(t)$ von X die Ableitung gerader Ordnung $\varphi^{(2l)}(0)$ besitzt (Cramér [2]).

Hinweis. Man beweise unter Verwendung symmetrischer Ableitungen die Beziehung

$$\begin{aligned}\varphi^{(2l)}(0) &= \lim_{h\to 0} \int_{-\infty}^{\infty} \left(\frac{e^{ihx} - e^{-ihx}}{2h}\right)^{2l} dF(x)\\ &= (-1)^l \lim_{h\to 0} \int_{-\infty}^{\infty} \left(\frac{\sin hx}{h}\right)^{2l} dF(x).\end{aligned}$$

9. a) Man zeige (Zygmund [1]), daß

$$\lim_{a\to\infty} \int_{-a}^{a} x\, dF(x) = \frac{\varphi'(0)}{i}$$

gilt, wenn die charakteristische Funktion $\varphi(t)$ der Verteilungsfunktion $F(x)$ an der Stelle 0 differenzierbar ist. Die linke Seite der obenstehenden Gleichung nennen wir den *verallgemeinerten Mittelwert* der Zufallsvariablen X.

b) Man zeige (PITMAN [2]), daß die charakteristische Funktion $\varphi(t)$ an der Stelle $t = 0$ differenzierbar ist, wenn der verallgemeinerte Mittelwert einer Zufallsvariablen X existiert und die Beziehung

$$\lim_{a\to\infty} aP(|X| > a) = 0$$

gilt.

c) Man zeige, daß

$$\varphi(t) = C \sum_{\substack{n=-\infty \\ n\neq 0, \pm 1, \pm 2,}}^{\infty} \frac{\cos nt}{n^2 \log |n|},$$

wobei C eine geeignet gewählte Konstante bedeutet, die charakteristische Funktion einer gewissen Zufallsvariablen ist, die keinen Erwartungswert besitzt; man zeige ferner, daß dann $\varphi(t)$ differenzierbar ist.

10. Man beweise den Satz 3.2.1.

Hinweis. Unter Verwendung der Ungleichung (3.4.5) zeige man, daß

$$\lim_{k\to\infty} \frac{\beta_k}{k!} r^k = 0$$

gilt. Anschließend zeige man, daß für $|t| < r$ die Reihe

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{m_k}{k!} (it)^k$$

konvergent ist. Schließlich setze man eine analytische Fortsetzung an, um zu zeigen, daß diese Reihe für beliebige t konvergiert.

11. Man zeige, daß die charakteristischen Funktionen $\varphi_1(t)$, $\varphi_2(t)$ und $\varphi_3(t)$ der Gleichung

$$\varphi_1(t)\,\varphi_2(t) = \varphi_1(t)\,\varphi_3(t)$$

genügen können, obwohl $\varphi_2(t)$ und $\varphi_3(t)$ nicht identisch gleich sind.

12. a) Man beweise den Satz (BOCHNER [1]): Ist x_0 Sprungstelle der Verteilungsfunktion $F(x)$ und ist $F(x_0 + 0) - F(x_0) = p$, so gilt

$$p = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itx_0}\,\varphi(t)\,dt,$$

wobei $\varphi(t)$ die charakteristische Funktion der Verteilungsfunktion $F(x)$ ist.

b) Man beweise (WIENER [2]), daß die Verteilungsfunktion $F(x)$ dann und nur dann stetig ist, wenn

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\varphi(t)|\,dt = 0$$

gilt.

c) Aus dem eben Bewiesenen ist die Beziehung (LORCH und NEWMAN [1])

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\varphi(t)|\, dt = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{itx_j} |\sum_j [F(x_j + 0) - F(x_j)]|\, dt$$

herzuleiten, wobei die x_j die Sprungstellen der Verteilungsfunktion $F(x)$ sind, wenn solche überhaupt existieren.

13. Man beweise (LÉVY [1]), daß für jede charakteristische Funktion $\varphi(t)$ die Beziehung

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\varphi(t)|^2\, dt = \sum_j \bigl(F(x_j + 0) - F(x_j)\bigr)$$

gilt, wobei die Summation sich über alle existierenden Sprungstellen x_j der Verteilungsfunktion $F(x)$ erstreckt, die der charakteristischen Funktion $\varphi(t)$ entspricht.

14. Man zeige: Genügt eine charakteristische Funktion $\varphi(t)$ für eine Folge von Punkten $t_k \neq 0$ $(k = 1, 2, \ldots)$, die gegen Null konvergiert, der Gleichung $|\varphi(t_k)| = 1$, so ist $|\varphi(t)| \equiv 1$ identisch erfüllt.

15. Man bezeichnet den Ausdruck

$$m_{[l]} = \sum_k k(k-1)\cdots(k-l+1)p_k,$$

wobei $p_k = P(X = k)$ und $\sum_k p_k = 1$ ist, als das *faktorielle Moment der Ordnung* l der ganzzahligen Zufallsvariablen X.

a) Man drücke $m_{[l]}$ mit Hilfe erzeugender Funktionen aus.

b) Man ermittle die Beziehungen, die zwischen den faktoriellen und den gewöhnlichen Momenten bestehen.

c) Welchen Ausdruck könnte man als das zentrale faktorielle Moment der Ordnung l bezeichnen?

16. Man beweise, daß die zugehörige Verteilungsfunktion Momente beliebiger Ordnung besitzt, wenn die Reihe auf der rechten Seite der Definitionsgleichung (4.7.1) der erzeugenden Funktion für ein gewisses $s_0 > 1$ konvergent ist.

17. Als *Momentenerzeugende* einer Zufallsvariablen X oder ihrer Verteilungsfunktion $F(x)$ bezeichnen wir den Ausdruck

$$g(t) = E(e^{tX}),$$

falls dieser endlich ist.

a) Man schreibe den Ausdruck für $g(t)$ für stetige und diskrete Zufallsvariable auf.

b) Man bestimme die Funktion $g(t)$ für die in den Beispielen 4.2.1 und 4.2.2 betrachteten Verteilungen.

c) Wie lassen sich die Momente mit Hilfe der Momentenerzeugenden ausdrücken? Vgl. hierzu auch die Aufgaben 5.14.27 und 5.14.28.

5. EINIGE WAHRSCHEINLICHKEITS-VERTEILUNGEN

5.1. Die Ein- und Zweipunktverteilungen

In den vorhergehenden Kapiteln haben wir verschiedene Wahrscheinlichkeitsverteilungen kennengelernt. In diesem Kapitel wollen wir uns eingehender mit einigen speziellen Wahrscheinlichkeitsverteilungen beschäftigen, welche große theoretische oder praktische Bedeutung besitzen.

Wir beginnen mit der Einpunktverteilung.

Definition 5.1.1. Die Zufallsvariable X hat eine *Einpunktverteilung*, wenn es einen Punkt x_0 gibt, für den

$$P(X = x_0) = 1 \tag{5.1.1}$$

ist.

Wir sagen auch, daß bei dieser Verteilung die ganze Wahrscheinlichkeitsmasse *in einem Punkt konzentriert* ist. Offensichtlich weist eine Zufallsvariable mit Einpunktverteilung eine ausgeartete Verteilung gemäß Definition 3.6.4 auf.

Die Formel (5.1.1) gibt die Wahrscheinlichkeitsfunktion an. Die Verteilungsfunktion ist hier durch die Bedingungen

$$F(x) = \begin{cases} 0 & \text{für} \quad x \leqq x_0, \\ 1 & \text{für} \quad x > x_0 \end{cases} \tag{5.1.2}$$

bestimmt. Die charakteristische Funktion dieser Verteilung erhalten wir aus (4.1.2):

$$\varphi(t) = e^{itx_0}. \tag{5.1.3}$$

Aus (4.2.4) erkennt man, daß

$$m_1 = x_0$$

und allgemein für jedes k

$$m_k = x_0^k$$

ist. Daraus erhalten wir

$$D^2(X) = m_2 - m_1^2 = x_0^2 - x_0^2 = 0.$$

Ist umgekehrt die Dispersion einer Zufallsvariablen X gleich Null, dann hat X eine Einpunktverteilung.

Nach Voraussetzung ist

$$D^2(X) = E[X - E(X)]^2 = 0. \tag{5.1.4}$$

Da der Ausdruck $[X - E(X)]^2$ niemals negativ ist, ist die Gleichung (5.1.4) nur dann erfüllt, wenn

$$P\big(X - E(X) = 0\big) = 1 \quad \text{oder} \quad P\big(X = E(X)\big) = 1$$

ist. Nach Formel (5.1.1) hat also die Zufallsvariable X eine Einpunktverteilung.

Definition 5.1.2. Die Zufallsvariable X besitzt eine *Zweipunktverteilung*, wenn sie nur zwei Werte x_1 und x_2 mit positiver Wahrscheinlichkeit annehmen kann. Die Wahrscheinlichkeitsfunktion wird dann durch

$$P(X = x_1) = p, \quad P(X = x_2) = 1 - p \quad (0 < p < 1) \tag{5.1.5}$$

gegeben.

Häufig betrachtet man den Fall $x_1 = 1$ und $x_2 = 0$. Statt der Formeln (5.1.5) haben wir dann

$$P(X = 1) = p, \quad P(X = 0) = 1 - p \quad (0 < p < 1). \tag{5.1.6}$$

Diese Verteilung wollen wir einfach *Null-Eins-Verteilung* nennen.

Die charakteristische Funktion der Verteilung (5.1.6) wird durch

$$\varphi(t) = pe^{it\cdot 1} + (1-p)e^{it\cdot 0} = pe^{it} + 1 - p = 1 + p(e^{it} - 1) \tag{5.1.7}$$

gegeben.

Aus (4.2.4) folgt für jedes k

$$m_k = p. \tag{5.1.8}$$

Daraus erhalten wir

$$D^2(X) = m_2 - m_1^2 = p - p^2 = p(1-p). \tag{5.1.9}$$

Die Formeln (3.2.9) und (3.2.17) ergeben

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 = p - 3p^2 + 2p^3 = p(1-p)(1-2p),$$

$$\gamma = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{p(1-p)(1-2p)}{p^{\frac{3}{2}}(1-p)^{\frac{3}{2}}} = \frac{1-2p}{\sqrt{p(1-p)}}. \tag{5.1.10}$$

Bei $p = 0{,}5$ wird $\gamma = 0$. Die Zufallsvariable X ist dann symmetrisch verteilt.

5.2. Das Bernoullische Versuchsschema. Die Binomialverteilung

Im Beispiel 3.1.3 haben wir die Binomialverteilung kennengelernt. Bei dem folgenden Versuchsschema, das nach BERNOULLI benannt wird, tritt eine Zufallsvariable X auf, die eine Binomialverteilung besitzt.

Wir machen n Versuche. Das Ergebnis jedes Versuchs ist zufällig und kann entweder das Ereignis A (mit der Wahrscheinlichkeit p) oder das zu A komplementäre Ereignis $\bar{A}$ (mit der Wahrscheinlichkeit $q = 1 - p$) sein. Die Ergebnisse der n Versuche sind unabhängig. Wir ordnen dem Ergebnis A die Zahl Eins und dem Ergebnis $\bar{A}$ die Zahl Null zu. Bei n Versuchen kann das Ereignis A genau k-mal $(k = 0, 1, 2, \ldots, n)$ eintreten. Die Anzahl des Eintretens des Ereignisses A bei n Versuchen ist eine Zufallsvariable X, die also die Werte $k = 0, 1, \ldots, n$ annehmen kann.

Im Beispiel 3.1.3 wurde gezeigt, daß die Wahrscheinlichkeitsfunktion dieser Zufallsvariablen durch

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{5.2.1}$$

gegeben wird. Dann ist die Verteilungsfunktion der Binomialverteilung gleich

$$F(x) = P(X < x) = \sum_{k<x} \binom{n}{k} p^k (1 - p)^{n-k},$$

wobei die Summation über alle nichtnegativen ganzen Zahlen, die kleiner als x sind, zu erstrecken ist.

Wie wir sehen, stimmt für $n = 1$ die Binomialverteilung mit der Null-Eins-Verteilung überein. Für $n \geqq 2$ können wir die Binomialverteilung folgendermaßen aus der Null-Eins-Verteilung erhalten:

Es seien X_r $(r = 1, 2, \ldots, n)$ unabhängige Zufallsvariable mit denselben Null-Eins-Verteilungen. Die Wahrscheinlichkeitsfunktion der X_r habe die Gestalt

$$P(X_r = 1) = p, \quad P(X_r = 0) = 1 - p.$$

Wir betrachten die Zufallsvariable, die gleich der Summe der unabhängigen Zufallsvariablen X_r ist:

$$X = X_1 + X_2 + \cdots + X_n. \tag{5.2.2}$$

X kann also die Werte $k = 0, 1, \ldots, n$ annehmen. Das Ereignis $X = k$ besteht darin, daß k der n Zufallsvariablen X_r den Wert Eins und $n - k$ Zufallsvariable den Wert Null annehmen. Für ein gegebenes k kann dieses Ereignis auf $\binom{n}{k}$ verschiedene Arten realisiert werden. Wenn wir die Unabhängigkeit der X_r beachten,

so erhalten wir (5.2.1). Aus den Formeln (5.1.7) und (5.2.2) finden wir als charakteristische Funktion $\varphi(t)$ der Zufallsvariablen X:

$$\varphi(t) = [1 + p(e^{it} - 1)]^n. \tag{5.2.3}$$

Aus (4.2.4) erhalten wir die Momente dieser Verteilung, insbesondere

$$m_1 = np, \quad m_2 = np + n(n-1)p^2, \tag{5.2.4}$$

$$\mu_2 = np(1-p), \quad \mu_3 = np(1-p)(1-2p).$$

Wir finden also

$$\gamma = \frac{1-2p}{\sqrt{np(1-p)}}. \tag{5.2.5}$$

Die Formel für μ_2 erhielten wir schon auf anderem Wege in Beispiel 3.2.3.

Beispiel 5.2.1. In Tabelle 5.2.1 sind die Binomialverteilungen für $n = 20$ und die p-Werte

$$p_1 = 0{,}1, \quad p_2 = 0{,}3, \quad p_3 = 0{,}5$$

angegeben. In der ersten Spalte der Tabelle sind die Werte $k = 0, 1, \ldots, 20$ aufgeführt, während in den übrigen Spalten die Wahrscheinlichkeiten angegeben sind, mit denen die Zufallsvariable diese Werte von k annimmt. Die Wahrscheinlichkeiten sind bis auf 0,0001 genau angegeben.

Tabelle 5.2.1.

k	$P(X = k)$			k	$P(X = k)$		
	$p_1 = 0{,}1$	$p_2 = 0{,}3$	$p_3 = 0{,}5$		$p_1 = 0{,}1$	$p_2 = 0{,}3$	$p_3 = 0{,}5$
0	0,1216	0,0008	—	11	—	0,0120	0,1602
1	0,2702	0,0068	—	12	—	0,0039	0,1201
2	0,2852	0,0278	0,0002	13	—	0,0010	0,0739
3	0,1901	0,0716	0,0011	14	—	0,0002	0,0370
4	0,0898	0,1304	0,0046	15	—	—	0,0148
5	0,0319	0,1789	0,0148	16	—	—	0,0046
6	0,0089	0,1916	0,0370	17	—	—	0,0011
7	0,0020	0,1643	0,0739	18	—	—	0,0002
8	0,0004	0,1144	0,1201	19	—	—	—
9	0,0001	0,0654	0,1602	20	—	—	—
10	—	0,0308	0,1762				

In Abb. 5.2.1 sind die Verteilungen dargestellt.

Man sieht: Je weniger die Wahrscheinlichkeit p von $\frac{1}{2}$ abweicht, desto symmetrischer ist die Verteilung; gleichzeitig wächst aber auch die Streuung. Diese Ergebnisse waren zu

erwarten, wir hätten nur die Werte der Parameter μ_2 und γ für die betrachteten p-Werte und $n = 20$ zu vergleichen brauchen. Diese aus (5.2.4) und (5.2.5) berechneten Parameterwerte sind in Tabelle 5.2.2 zusammengestellt.

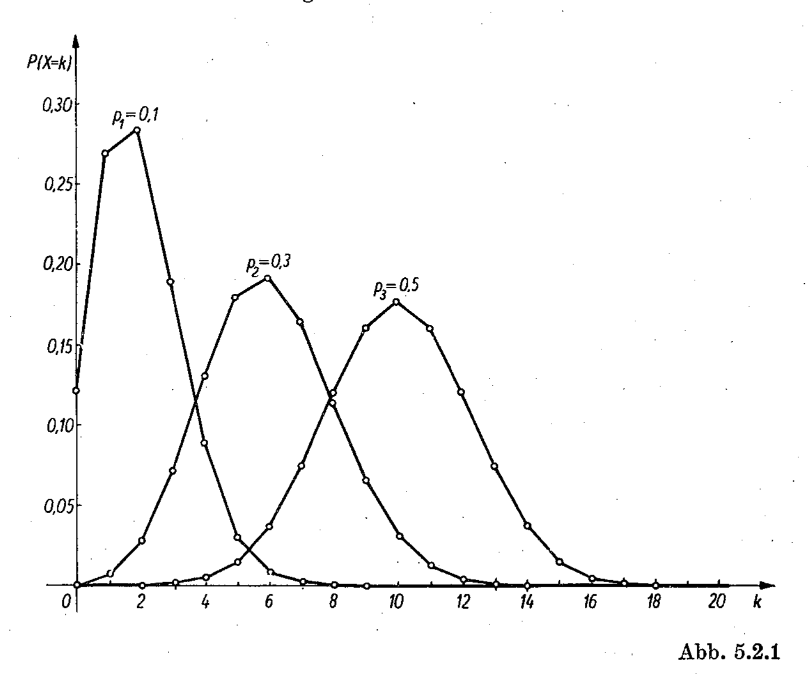


Abb. 5.2.1

Tabelle 5.2.2.

	$p_1 = 0{,}1$	$p_2 = 0{,}3$	$p_3 = 0{,}5$
$\sigma = \sqrt{\mu_2}$	1,34	2,05	2,24
γ	0,597	0,195	0,000

Es seien X und Y zwei binomial verteilte, unabhängige Zufallsvariable, und es mögen die charakteristischen Funktionen der Variablen X und Y die Gestalt

$$\varphi_1(t) = [1 + p(e^{it} - 1)]^{n_1},$$

$$\varphi_2(t) = [1 + p(e^{it} - 1)]^{n_2}$$

haben. Wir betrachten die Variable

$$Z = X + Y.$$

Wegen der Unabhängigkeit von X und Y wird die charakteristische Funktion

von Z durch

$$\varphi(t) = [1 + p(e^{it} - 1)]^{n_1+n_2} \tag{5.2.6}$$

gegeben.

Wie man aus (5.2.6) erkennt, ist die Variable Z ebenfalls binomial verteilt. Sie ist gleich der Summe von $n_1 + n_2$ unabhängigen Summanden, die alle dieselbe Null-Eins-Verteilung haben. Das ist der sogenannte *Additionssatz* für die Binomialverteilung.

In den Anwendungen untersucht man häufig die Verteilung der Zufallsvariablen

$$Y = \frac{X}{n},$$

wenn X binomial verteilt ist. Die Zufallsvariable Y kann die Werte

$$\frac{k}{n} = 0, \quad \frac{1}{n}, \ldots, \quad \frac{n-1}{n}, \quad 1$$

annehmen. Da die Wahrscheinlichkeit dafür, daß $Y = \frac{k}{n}$ gilt, gleich der Wahrscheinlichkeit dafür ist, daß $X = k$ gilt, wird die Wahrscheinlichkeitsfunktion der Zufallsvariablen Y durch die Formel (5.2.1) gegeben:

$$P\left(Y = \frac{k}{n}\right) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Unter Berücksichtigung der Formeln (4.2.15) und (5.2.3) erhalten wir für die charakteristische Funktion der Zufallsvariablen Y:

$$\varphi(t) = \left[1 + p\left(e^{\frac{it}{n}} - 1\right)\right]^n. \tag{5.2.7}$$

Die Momente erhalten wir aus der Formel (4.2.4); insbesondere ist

$$m_1 = p, \quad m_2 = \frac{p}{n} + \frac{n-1}{n} p^2, \quad \mu_2 = \frac{p(1-p)}{n}. \tag{5.2.8}$$

5.3. Das Poissonsche Versuchsschema. Die verallgemeinerte Binomialverteilung

Poisson betrachtete das folgende Versuchsschema: Man macht n Versuche. Als Ergebnis des k-ten Versuchs $(k = 1, 2, \ldots, n)$ kann das Ereignis A mit der Wahrscheinlichkeit p_k eintreten. Die Wahrscheinlichkeit des zu A komplementären Ereignisses $\bar{A}$ ist also gleich $q_k = 1 - p_k$. Die Ergebnisse der n Versuche sind unabhängig. Im Gegensatz zum Bernoullischen Schema brauchen hier in den ein-

zelnen Versuchen die Wahrscheinlichkeiten für das Eintreten des Ereignisses A nicht einander gleich zu sein.

Wir ordnen dem Ereignis A die Zahl Eins und dem Ereignis $\bar{A}$ die Zahl Null zu. Die Anzahl der Fälle, in denen das Ereignis A bei n Versuchen eintritt, ist eine zufällige Veränderliche. Wir sagen, daß diese Zufallsvariable eine *verallgemeinerte Binomialverteilung* besitzt.

Eine Zufallsvariable Z, die eine verallgemeinerte Binomialverteilung besitzt, kann auch als Summe

$$Z = Z_1 + \cdots + Z_n \tag{5.3.1}$$

dargestellt werden, wobei die Zufallsvariablen Z_k $(k = 1, \ldots, n)$ unabhängig sind und Null-Eins-Verteilungen mit den Wahrscheinlichkeitsfunktionen

$$P(Z_k = 1) = p_k, \quad P(Z_k = 0) = 1 - p_k$$

haben.

Die Formel für die Wahrscheinlichkeitsfunktion der Zufallsvariablen Z ist nicht so einfach wie die entsprechende Formel für die Binomialverteilung. Die Wahrscheinlichkeiten dafür, daß $Z = r$, finden wir, indem wir die Wahrscheinlichkeiten aller möglichen Anordnungen, die aus r Einsen und $n - r$ Nullen bestehen, addieren.

Beispiel 5.3.1. Wir betrachten drei Ladungen Apfelsinen. Der Anteil der verfaulten („fehlerhaften") Apfelsinen betrage in der ersten Ladung $p_1 = 0{,}02$, in der zweiten $p_2 = 0{,}05$, in der dritten $p_3 = 0{,}01$. Aus jeder Ladung greifen wir aufs Geratewohl eine Apfelsine heraus. Ergreifen wir eine gute Apfelsine, so schreiben wir die Zahl Eins hin, ergreifen wir dagegen eine verfaulte, die Zahl Null. Z_1, Z_2, Z_3 sind hier Zufallsvariable, die die Werte Eins bzw. Null annehmen, wenn wir aus der ersten, zweiten oder dritten Ladung eine gute bzw. eine schlechte Apfelsine herausgreifen. Diese Veränderlichen sind unabhängig, und es ist

$$P(Z_1 = 1) = 0{,}98, \quad P(Z_2 = 1) = 0{,}95, \quad P(Z_3 = 1) = 0{,}99.$$

Wir betrachten die Zufallsvariable $Z = Z_1 + Z_2 + Z_3$. Sie kann die Werte $r = 0, 1, 2, 3$ annehmen, und wir berechnen der Reihe nach die Wahrscheinlichkeiten dafür, daß $Z = r$ ist, wobei wir die Unabhängigkeit der Z_1, Z_2, Z_3 benutzen:

$$P(Z = 0) = P(Z_1 = 0)P(Z_2 = 0)P(Z_3 = 0) = 0{,}00001,$$

$$\begin{aligned} P(Z = 1) &= P(Z_1 = 0)P(Z_2 = 0)P(Z_3 = 1) \\ &\quad + P(Z_1 = 0)P(Z_2 = 1)P(Z_3 = 0) + P(Z_1 = 1)P(Z_2 = 0)P(Z_3 = 0) \\ &= 0{,}00167, \end{aligned}$$

$$\begin{aligned} P(Z = 2) &= P(Z_1 = 0)P(Z_2 = 1)P(Z_3 = 1) \\ &\quad + P(Z_1 = 1)P(Z_2 = 0)P(Z_3 = 1) + P(Z_1 = 1)P(Z_2 = 1)P(Z_3 = 0) \\ &= 0{,}07663, \end{aligned}$$

$$P(Z = 3) = P(Z_1 = 1)P(Z_2 = 1)\,P(Z_3 = 1) = 0{,}92169.$$

Hieraus folgt sofort $P(Z = 0) + P(Z = 1) + P(Z = 2) + P(Z = 3) = 1.$

Die Zufallsvariable Z, die durch die Formel (5.3.1) definiert ist, habe eine verallgemeinerte Binomialverteilung. Ihre charakteristische Funktion erhalten wir dann aus (5.1.7), wobei wir beachten, daß die Summanden Z_k unabhängig sind:

$$\varphi(t) = \prod_{k=1}^{n} [1 + p_k(e^{it} - 1)]. \tag{5.3.2}$$

Um mit Hilfe der Formel (5.3.2) die ersten zwei Momente der Zufallsvariablen Z berechnen zu können, benutzen wir die Formel (4.2.4). Wir erhalten

$$m_1 = \sum_{k=1}^{n} p_k, \quad m_2 = \sum_{k=1}^{n} p_k + \sum_{\substack{l=1 \\ l \neq k}}^{n} \sum_{k=1}^{n} p_l p_k, \tag{5.3.3}$$

$$\mu_2 = \sum_{k=1}^{n} p_k(1 - p_k).$$

Wie wir sehen, sind die Formeln (5.2.4) für m_1, m_2 und μ_2 Spezialfälle der entsprechenden Formeln (5.3.3).

Beispiel 5.3.2. Für die Zufallsvariable Z aus dem Beispiel 5.3.1 berechnen wir den Mittelwert und die Standardabweichung.

Wir erhalten

$$E(Z) = m_1 = (1 - p_1) + (1 - p_2) + (1 - p_3) = 2{,}92,$$

$$\sigma = \sqrt{\mu_2} = \sqrt{0{,}0196 + 0{,}0475 + 0{,}0099} = \sqrt{0{,}0770} \approx 0{,}28.$$

5.4. Die Pólyasche und die hypergeometrische Verteilung

A. In der Praxis trifft man häufig Verteilungen an, die sich in das sogenannte *Pólyasche Schema* bringen lassen.

Stellen wir uns Folgendes vor: In einer Urne befinden sich b weiße und c schwarze Kugeln. Wir setzen $b + c = N$. Aus der Urne nehmen wir aufs Geratewohl eine Kugel heraus. Bevor wir die nächste Kugel nehmen, legen wir diese Kugel zurück und legen außerdem noch s Kugeln derselben Farbe wie die herausgegriffene in die Urne hinein. Das wiederholen wir n-mal. Wir bezeichnen mit X die Zufallsvariable, die den Wert k ($k = 0, 1, \ldots, n$) annimmt, wenn bei n Ziehungen k-mal weiße Kugeln erscheinen, und wollen die Wahrscheinlichkeitsfunktion von X bestimmen.

Die Wahrscheinlichkeit dafür, daß k-mal nacheinander weiße Kugeln gezogen werden, ist gleich

$$\frac{b(b + s)\cdots[b + (k - 1)s]}{N(N + s)\cdots[N + (k - 1)s]}.$$

Ebenso ist die Wahrscheinlichkeit dafür, daß zuerst k-mal nacheinander weiße

Kugeln und dann $(n-k)$-mal nacheinander schwarze Kugeln gezogen werden, gleich

$$\frac{b(b+s)\cdots[b+(k-1)s]c(c+s)\cdots[c+(n-k-1)s]}{N(N+s)\cdots[N+(n-1)s]}.$$

Der letzte Ausdruck ist ebenfalls gleich der Wahrscheinlichkeit dafür, daß man k weiße und $n-k$ schwarze Kugeln in irgendeiner Reihenfolge zieht. Die Reihenfolge des Ziehens von weißen und schwarzen Kugeln hat nämlich nur Einfluß auf die Faktorenanordnung des Zählers. Da man k weiße und $n-k$ schwarze Kugeln auf $\binom{n}{k}$ verschiedene Arten anordnen (ziehen) kann, ergibt sich

$$P(X=k)=\binom{n}{k}\frac{b(b+s)\cdots[b+(k-1)s]c(c+s)\cdots[c+(n-k-1)s]}{N(N+s)\cdots[N+(n-1)s]}. \tag{5.4.1}$$

Definition 5.4.1. Wir sagen, daß die Zufallsvariable X mit der durch die Formel (5.4.1) gegebenen Wahrscheinlichkeitsfunktion eine *Pólyasche Verteilung* hat.

Es sei nun

$$Np=b, \quad Nq=c, \quad Na=s.$$

Wie wir sehen, sind p bzw. q die Wahrscheinlichkeiten dafür, beim ersten Mal eine weiße bzw. schwarze Kugel zu erhalten. Die Formel (5.4.1) nimmt dann die Gestalt

$$P(X=k)=\binom{n}{k}\frac{p(p+a)\cdots[p+(k-1)a]q(q+a)\cdots[q+(n-k-1)a]}{1\cdot(1+a)\cdots[1+(n-1)a]} \tag{5.4.2}$$

an. Offenbar ist

$$\sum_{k=0}^{n}\binom{n}{k}\frac{p(p+a)\cdots[p+(k-1)a]q(q+a)\cdots[q+(n-k-1)a]}{1\cdot(1+a)\cdots[1+(n-1)a]}=1. \tag{5.4.3}$$

Wir berechnen die Momente erster und zweiter Ordnung. Das erste Moment wird durch die Formel

$$\begin{aligned}E(X)&=\sum_{k=0}^{n}kP(X=k)\\&=pn\sum_{k=1}^{n}\binom{n-1}{k-1}\frac{(p+a)\cdots[p+(k-1)a]q(q+a)\cdots[q+(n-k-1)a]}{(1+a)\cdots[1+(n-1)a]}\end{aligned}$$

gegeben. Für $l=k-1$ erhalten wir

$$E(X)=pn\sum_{l=0}^{n-1}\binom{n-1}{l}\frac{(p+a)\cdots(p+la)q(q+a)\cdots[q+(n-l-2)a]}{(1+a)\cdots[1+(n-1)a]}, \tag{5.4.4}$$

Wie man leicht bestätigt, ist der Ausdruck unter dem Summenzeichen in der letzten Formel die Wahrscheinlichkeit dafür, bei $n-1$ Ziehungen l-mal weiße und $(n-l-1)$-mal schwarze Kugeln zu erhalten, wenn das Ziehen der Kugeln nach dem Pólyaschen Schema erfolgt und die Urne zu Beginn $b+s$ weiße und c schwarze Kugeln enthält, im ganzen also $N+s$ Kugeln. Wegen der Beziehung (5.4.3) folgt daraus, daß die Summe auf der rechten Seite der Formel (5.4.4) gleich Eins ist, also

$$E(X) = np. \tag{5.4.5}$$

Wir wollen jetzt das zweite Moment berechnen:

$$E(X^2) = \sum_{k=0}^{n} k^2 P(X=k)$$
$$= np \sum_{k=1}^{n} k \binom{n-1}{k-1} \frac{(p+a)\cdots[p+(k-1)a]q(q+a)\cdots[q+(n-k-1)a]}{(1+a)\cdots[1+(n-1)a]}.$$

Wie vorhin setzen wir $l = k-1$:

$$E(X^2) = np \sum_{l=1}^{n-1} (l+1) \binom{n-1}{l}$$
$$\times \frac{(p+a)\cdots(p+la)q(q+a)\cdots[q+(n-l-2)a]}{(1+a)\cdots[1+(n-1)a]}$$
$$= np \left\{ \sum_{l=1}^{n-1} l \binom{n-1}{l} \frac{(p+a)\cdots(p+la)q(q+a)\cdots[q+(n-l-2)a]}{(1+a)\cdots[1+(n-1)a]} \right.$$
$$\left. + \sum_{l=0}^{n-1} \binom{n-1}{l} \frac{(p+a)\cdots(p+la)q(q+a)\cdots[q+(n-l-2)a]}{(1+a)\cdots[1+(n-1)a]} \right\}$$
$$= np(A+B).$$

Nach einfachen Umformungen erhalten wir

$$A = \frac{(p+a)(n-1)}{1+a} \sum_{r=0}^{n-2} \binom{n-2}{r}$$
$$\times \frac{(p+2a)\cdots[p+(r+1)a]q\cdots[q+(n-r-3)a]}{(1+2a)\cdots[1+(n-1)a]}, \tag{5.4.6}$$

$$B = \sum_{l=0}^{n-1} \binom{n-1}{l} \frac{(p+a)\cdots(p+la)q(q+a)\cdots[q+(n-l-2)a]}{(1+a)\cdots[1+(n-1)a]}. \tag{5.4.7}$$

Der Ausdruck B ist identisch mit der Summe in der Formel (5.4.4), also ist $B = 1$. Wir bemerken weiter, daß der Ausdruck unter dem Summenzeichen in der Formel (5.4.6) gleich der Wahrscheinlichkeit dafür ist, daß man bei $n-2$ Ziehungen nach dem Pólyaschen Schema r-mal weiße und $(n-r-2)$-mal

schwarze Kugeln erhält, wenn die Urne zu Beginn $b + 2s$ weiße und c schwarze, also insgesamt $N + 2s$ Kugeln enthält. Aus Formel (5.4.3) folgt also

$$A = \frac{(p + a)(n - 1)}{1 + a}.$$

Wir erhalten schließlich

$$E(X^2) = np\left[\frac{(p + a)(n - 1)}{1 + a} + 1\right] = np \cdot \frac{np + q + na}{1 + a}. \tag{5.4.8}$$

Unter Beachtung der Formel (5.4.5) erhalten wir

$$D^2(X) = npq\,\frac{1 + na}{1 + a}. \tag{5.4.9}$$

Das Pólyasche Schema findet Anwendung bei Erscheinungen, die sich wie ansteckende Krankheiten verhalten, bei denen also das Eintreffen eines Ereignisses (lies: eine Erkrankung) die Wahrscheinlichkeit der Ansteckung erhöht.

Im Pólyaschen Schema kann s auch negativ sein, denn da die Ungleichungen

$$b + (k - 1)s \geqq 1 \quad \text{sowie} \quad c + (n - k - 1)s \geqq 1$$

gelten müssen, genügt k der Doppelungleichung

$$\max\left(0, n - 1 + \frac{c - 1}{s}\right) \leqq k \leqq \min\left(n, \frac{1 - b}{s} + 1\right).$$

B. Die Größen N, b und c sollen so gegen Unendlich streben, daß

$$p = \frac{b}{N} = \text{const} \tag{5.4.10}$$

bleibt. Dann ist $q = 1 - p$ ebenfalls konstant. Wir setzen außerdem noch voraus, daß $\lim\limits_{N\to\infty} a = 0$ ist. Das ist insbesondere dann erfüllt, wenn s konstant bleibt und N unendlich wird. Aus den Formeln (5.4.1) und (5.4.2) folgt dann

$$\lim_{N\to\infty} P(X = k) = \binom{n}{k} p^k q^{n-k}. \tag{5.4.11}$$

Wir erhielten also:

Satz 5.4.1. *Ist für* $N = 1, 2, \ldots$ *die Gleichung* (5.4.10) *erfüllt und ist* $\lim\limits_{N\to\infty} a = 0$, *dann konvergiert die Wahrscheinlichkeitsfunktion einer Zufallsvariablen* X, *die eine Pólyasche Verteilung besitzt, gegen die Wahrscheinlichkeitsfunktion der Binomialverteilung.*

C. Ein Spezialfall der Pólyaschen Verteilung ist die *hypergeometrische Verteilung.* Bei dieser Verteilung ist $s = -1$, wir geben also einfach die gezogene Kugel vor

der Wahl der nächsten nicht zurück. Die Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung erhalten wir aus der Formel (5.4.2), wenn wir $a = -\frac{1}{N}$ setzen. Für Werte von k, die der Ungleichung

$$\max(0, n - Nq) \leqq k \leqq \min(n, Np)$$

genügen, finden wir

$$P(X = k) = \binom{n}{k} \frac{Np(Np-1)\cdots(Np-k+1)Nq\cdots(Nq-n+k+1)}{N(N-1)\cdots(N-n+1)}$$

$$= \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}. \tag{5.4.12}$$

Der Mittelwert $E(X)$ beträgt hier np, und die Formel für die Dispersion nimmt folgende Gestalt an:

$$D^2(X) = \frac{N-n}{N-1} npq. \tag{5.4.13}$$

Die hypergeometrische Verteilung findet Anwendung bei der Qualitätskontrolle der Massenproduktion. Es bestehe z. B. eine Warenladung aus b guten und $N - b = c$ fehlerhaften Exemplaren. Das gute Exemplar spielt hier die Rolle einer weißen, das fehlerhafte die Rolle einer schwarzen Kugel, an die Stelle der Urne tritt die Warenladung. Zur Qualitätskontrolle greifen wir aus der Ladung n Exemplare aufs Geratewohl heraus und legen vor der Wahl des nächsten das eben gewählte Exemplar nicht in die Ladung zurück. Wären die Größen b und c bekannt, dann könnte man die oben erhaltenen Formeln anwenden und die Wahrscheinlichkeit dafür berechnen, daß sich unter den n herausgegriffenen Exemplaren k gute befinden. In der Praxis sind jedoch die Zahlen b und c nicht bekannt, und die Untersuchung der Qualität einer gewissen Anzahl von Exemplaren dient eben zur Abschätzung dieser unbekannten Zahlen. Mit Problemen dieser Art werden wir uns im zweiten Teil des Buches beschäftigen.

5.5. Die Poissonsche Verteilung

In Beispiel 4.2.1 haben wir eine Zufallsvariable X mit einer Poissonschen Verteilung betrachtet. Wir stellen nun die wichtigsten Eigenschaften einer solchen Zufallsvariablen zusammen.

Sie kann die Werte $r = 0, 1, 2, \ldots$ annehmen. Ihre Wahrscheinlichkeitsfunktion ist durch

$$P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda} \tag{5.5.1}$$

bestimmt, wobei λ eine positive Konstante ist. Die charakteristische Funktion dieser Zufallsvariablen hat nach Formel (4.2.6) die Gestalt

$$\varphi(t) = e^{\lambda(e^{it}-1)}.$$

Die Formeln (4.2.7) bis (4.2.9) ergeben

$$m_1 = \lambda, \quad m_2 = \lambda(\lambda + 1), \quad \mu_2 = \lambda.$$

Die Wahrscheinlichkeitsfunktion (5.5.1) kann man als Limes von Wahrscheinlichkeitsfunktionen der Binomialverteilung erhalten. Um das zu zeigen, beweisen wir den sogenannten *Poissonschen Satz* (vgl. POISSON [1]):

Satz 5.5.1. *Die Zufallsvariable X_n habe eine durch die Formel*

$$P(X_n = r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \tag{5.5.2}$$

gegebene Binomialverteilung, wobei r die Werte $0, 1, \ldots, n$ annehmen kann. Gilt für $n = 1, 2, 3, \ldots$ die Beziehung[1])

$$p = \frac{\lambda}{n}, \tag{5.5.3}$$

wobei $\lambda > 0$ eine Konstante ist, dann ist

$$\lim_{n\to\infty} P(X_n = r) = \frac{\lambda^r}{r!} e^{-\lambda}. \tag{5.5.4}$$

Da der Mittelwert der Binomialverteilung gleich np ist, besagt die Bedingung (5.5.3), daß der Mittelwert der Binomialverteilung sich bei wachsendem n nicht ändert.

Beweis. Wir formen (5.5.2) folgendermaßen um:

$$\begin{aligned} P(X_n = r) &= \frac{n!}{r!(n-r)!} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} \\ &= \frac{\lambda}{r!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1)\cdots(n-r+1)}{n^r} \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^r} \\ &= \frac{\lambda^r}{r!} \left(1 - \frac{\lambda}{n}\right)^n \frac{1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{r-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^r}. \end{aligned}$$

[1]) Der Satz gilt auch dann noch, wenn wir die Relation (5.5.3) durch $\lim_{n\to\infty} np = \lambda$ ersetzen.

Unter Berücksichtigung von

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda} \quad \text{sowie} \quad \lim_{n\to\infty}\frac{1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^r} = 1$$

erhalten wir die Beziehung (5.5.4).

Auf den Satz 5.5.1 kommen wir noch in 6.9.C zurück.

Abb. 5.5.1 stellt die Wahrscheinlichkeitsfunktion der Binomialverteilung für $n = 5$ und $p = 0{,}3$, also $\lambda = np = 1{,}5$, und die Wahrscheinlichkeitsfunktion der Poissonverteilung mit demselben Mittelwert $\lambda = 1{,}5$ dar. In Abb. 5.5.2 sind graphische Darstellungen der gleichen Wahrscheinlichkeitsfunktion für $n = 10$ und $p = 0{,}15$, also wieder für $\lambda = 1{,}5$, angegeben. Wäre n noch größer (z. B. gleich 100), dann würden die graphischen Darstellungen der Binomial- und der Poissonverteilung fast vollständig übereinstimmen.

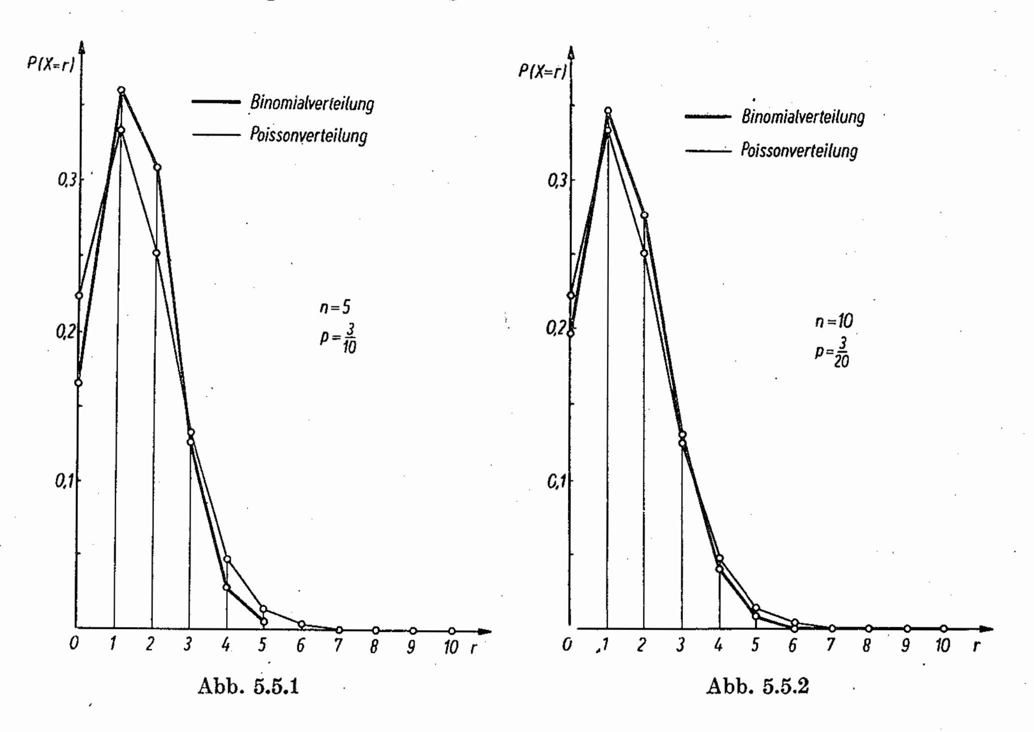


Abb. 5.5.1 Abb. 5.5.2

Häufig deutet man die Poissonsche Verteilung als die Verteilung einer Zufallsvariablen, die wohl viele verschiedene Werte (die Anzahl n ist groß), aber mit kleinen Wahrscheinlichkeiten (die Wahrscheinlichkeit $p = \frac{\lambda}{n}$ ist klein) annimmt.

Deshalb nennt man manchmal die Poissonsche Verteilung das *Gesetz der kleinen Zahlen*; diese Bezeichnung ist jedoch, wie zwei Beispiele zeigen werden, unbegründet.

BORTKIEWICZ [1], der sich mit der Untersuchung der Poissonschen Verteilung befaßte, bringt einige empirische Beispiele von zufälligen Ereignissen, auf die diese Verteilung angewandt werden kann. Wir zitieren eines davon:

Beispiel 5.5.1. Aus den Angaben der preußischen Armee hat BORTKIEWICZ die Anzahl der Soldaten in 10 Kavallerieregimentern berechnet, die in einem Zeitraum von 20 aufeinanderfolgenden Jahren infolge eines Huftritts starben.

Als zufälliges Ereignis betrachten wir hier, daß in einem Regiment im Laufe eines Jahres r Leute durch Huftritte umkamen ($r = 0, 1, 2, \ldots$). Es liegen $10 \cdot 20 = 200$ Beobachtungen vor, da 10 Regimenter im Laufe von 20 Jahren beobachtet wurden.

BORTKIEWICZ beobachtete nun die in Tabelle 5.5.1 angegebenen Häufigkeiten der Werte von r.

Tabelle 5.5.1. *Die Häufigkeit der Todesfälle an den Folgen eines Huftritts*

r	0	1	2	3	4
Häufigkeit	0,545	0,325	0,110	0,015	0,005
Wahrscheinlichkeit	0,544	0,331	0,101	0,021	0,003

Aus der mittleren Zeile dieser Tabelle berechnen wir den Mittelwert

$$E(X) = 0 \cdot 0{,}545 + 1 \cdot 0{,}325 + 2 \cdot 0{,}110 + 3 \cdot 0{,}015 + 4 \cdot 0{,}005 = 0{,}61.$$

Wir berechnen die entsprechenden Wahrscheinlichkeiten $P(X = r)$ für die Poissonverteilung mit $\lambda = 0{,}61$. (Diese Wahrscheinlichkeiten liest man gewöhnlich aus den Tafeln für die Poissonverteilung ab. Als Beispiel wollen wir aber die Wahrscheinlichkeiten unmittelbar berechnen.) Wir erhalten

$$P(X = 0) = e^{-0,61} = 0{,}544,$$

$$P(X = 1) = 0{,}61\, e^{-0,61} = 0{,}331,$$

$$P(X = 2) = \frac{0{,}61^2 e^{-0,61}}{2!} = 0{,}101,$$

$$P(X = 3) = \frac{0{,}61^3 e^{-0,61}}{3!} = 0{,}021,$$

$$P(X = 4) = \frac{0{,}61^4 e^{-0,61}}{4!} = 0{,}003.$$

Diese Werte sind in der unteren Zeile der Tabelle 5.5.1 aufgeführt. Wie wir sehen, unterscheiden sich die Wahrscheinlichkeiten nur wenig von den entsprechenden Häufigkeiten. Zu diesem Beispiel kehren wir noch einmal zurück (Beispiel 12.4.3).

In vielen physikalischen und technischen Problemen hat man es mit Verteilungen zu tun, die Annäherungen an die Poissonsche Verteilung sind. Wir bringen hier

ein Beispiel aus der Physik. In Kapitel 8 werden wir uns näher mit der Deutung des Mechanismus der zufälligen Erscheinungen, die der Poissonschen Verteilung unterliegen, befassen.

Beispiel 5.5.2. Wir zitieren hier die Ergebnisse der berühmten Rutherfordschen und Geigerschen Versuche. Die Physiker RUTHERFORD und GEIGER beobachteten die Anzahl der α-Teilchen, die von radioaktiven Substanzen in $n = 2608$ Zeitabschnitten von 7,5 Sekunden emittiert wurden.

In Tabelle 5.5.2 bedeutet n_i die Zahl der Zeitabschnitte, in der die Anzahl der emittierten α-Teilchen i betrug. Die durchschnittliche Anzahl λ der emittierten α-Teilchen in einem Zeitabschnitt von 7,5 Sekunden beträgt

$$\lambda = \frac{\sum n_i i}{n} = 3{,}87.$$

Wir schreiben

$$p_i = \frac{\lambda^i}{i!} e^{-\lambda}$$

und berechnen die Werte np_i. Der Leser beachte die gute Übereinstimmung der zweiten Spalte mit der dritten in Tabelle 5.5.2.

Tabelle 5.5.2.

i	n_i	np_i
0	57	54,399
1	203	210,523
2	383	407,361
3	525	525,496
4	532	508,418
5	408	393,515
6	273	253,817
7	139	140,325
8	45	67,882
9	27	29,189
10	16	17,075
	2608	2608,000

Ebenso wie für die Binomialverteilung kann man für unabhängige Zufallsvariable mit Poissonverteilungen einen Additionssatz beweisen.

Die unabhängigen Zufallsvariablen X_1 und X_2 mögen die entsprechenden Poissonschen Verteilungen

$$P(X_1 = r) = \frac{\lambda_1^r}{r!} e^{-\lambda_1},$$

$$P(X_2 = r) = \frac{\lambda_2^r}{r!} e^{-\lambda}$$

($r = 0, 1, 2, \ldots$) haben. Wir betrachten die Summe

$$X = X_1 + X_2.$$

Laut Formel (4.2.6) haben die charakteristischen Funktionen der Zufallsvariablen X_1 und X_2 die Form

$$\varphi_1(t) = e^{\lambda_1(e^{it}-1)},$$

$$\varphi_2(t) = e^{\lambda_2(e^{it}-1)}.$$

Auf Grund der Unabhängigkeit der Zufallsvariablen X_1 und X_2 ist die charakteristische Funktion der Zufallsvariablen X gleich

$$\varphi(t) = e^{(\lambda_1+\lambda_2)(e^{it}-1)}. \tag{5.5.5}$$

Die Formel (5.5.5) gibt die charakteristische Funktion einer nach Poisson verteilten Zufallsvariablen mit dem Mittelwert $\lambda_1 + \lambda_2$ an. Das ist der *Additionssatz* für unabhängige Zufallsvariable, die Poissonverteilungen besitzen.

Raikow [1] hat gezeigt, daß auch die Umkehrung des Satzes gilt: *Sind die Zufallsvariablen X_1 und X_2 unabhängig und hat $X = X_1 + X_2$ eine Poissonsche Verteilung, so weist auch jede der Variablen X_1 und X_2 eine Poissonsche Verteilung auf.*

Der Satz von Raikow gilt für beliebig viele unabhängige Zufallsvariable $X_1, X_2, \ldots, X_n$.

5.6. Die Rechtecksverteilung

A. Die einfachste stetige Verteilung ist die Rechtecksverteilung. In Beispiel 4.1.2 haben wir einen Spezialfall der Rechtecksverteilung betrachtet. Die allgemeine Definition ist folgende:

Definition 5.6.1. Die Zufallsvariable X hat eine *Rechtecksverteilung* oder ist *gleichverteilt*, wenn ihre Dichte $f(x)$ folgendermaßen gegeben ist:

$$f(x) = \begin{cases} \dfrac{1}{2h} & \text{für } a - h \leqq x \leqq a + h, \quad \text{wobei } a \text{ und } h\,(h > 0) \text{ gewisse Konstante sind,} \\ 0 & \text{sonst.} \end{cases} \tag{5.6.1}$$

Die Verteilungsfunktion $F(x)$ dieser Zufallsvariablen X ist dann gleich

$$F(x) = \begin{cases} 0 & \text{für } x < a - h, \\ \dfrac{1}{2h}\displaystyle\int_{a-h}^{x} dx = \dfrac{x-(a-h)}{2h} & \text{für } a - h \leqq x \leqq a + h, \\ 1 & \text{für } x > a + h. \end{cases} \tag{5.6.2}$$

Die charakteristische Funktion der Zufallsvariablen X wird durch die Formel

$$\varphi(t) = \frac{1}{2h}\int_{a-h}^{a+h} e^{itx}\,dx = \frac{1}{2h}\left[\frac{e^{itx}}{it}\right]_{a-h}^{a+h} = \frac{1}{2h}\,\frac{e^{it(a+h)} - e^{it(a-h)}}{it}$$

$$= e^{ita}\,\frac{\sin th}{th} \tag{5.6.3}$$

gegeben. Die Momente erhalten wir unmittelbar aus der Formel

$$m_k = \frac{1}{2h}\int_{a-h}^{a+h} x^k dx = \frac{1}{2h}\,\frac{(a+h)^{k+1} - (a-h)^{k+1}}{k+1}. \tag{5.6.4}$$

Insbesondere finden wir

$$m_1 = a, \quad m_2 = \frac{3a^2 + h^2}{3}.$$

Daraus ergibt sich

$$\mu_2 = m_2 - m_1^2 = \frac{h^2}{3}. \tag{5.6.5}$$

Wir können die Zufallsvariable X so linear transformieren, daß die transformierte Zufallsvariable Y im Intervall $[0,1]$ gleichverteilt ist. Es sei

$$Y = \frac{X - (a-h)}{2h}.$$

Die Dichte $f_1(y)$ der Zufallsvariablen Y ist dann in folgender Weise bestimmt:

$$f_1(y) = \begin{cases} 1 & \text{im Intervall } [0,1], \\ 0 & \text{sonst.} \end{cases} \tag{5.6.6}$$

Diese Rechtecksverteilung haben wir schon in 4.1.2 betrachtet. Die Dichte der Verteilung ist in Abb. 5.6.1 dargestellt.

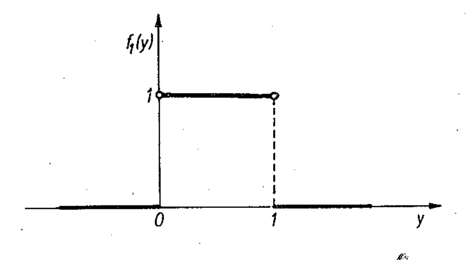


Abb. 5.6.1

B. Bei statistischen Problemen stößt man häufig auf Rechtecksverteilungen. Bemerkenswert ist die Tatsache, daß die Zufallsvariable $Y = F(X)$ im Intervall $[0,1]$ gleichverteilt ist, wenn die Verteilungsfunktion $F(x)$ der Zufallsvariablen X stetig ist.

In der Tat, jedem unendlichen Werteintervall $-\infty < X \leqq x$ der Zufallsvariablen X entspricht eine Wertemenge der Zufallsvariablen Y, die im Intervall $0 \leqq Y \leqq y = F(x)$ enthalten ist.

Andererseits entspricht auf Grund der Voraussetzung, daß die Verteilungsfunktion $F(x)$ stetig sein soll, jedem Wert $0 \leqq y \leqq 1$ mindestens ein Wert von x, der die Beziehung

$$y = F(x) = P(X < x) \tag{5.6.7}$$

erfüllt. Die Transformation (5.6.7) braucht aber nicht eineindeutig zu sein, da die Urbilder $F^{-1}(y)$ gewisser y Intervalle sein können, in denen die Verteilungsfunktion $F(x)$ konstant ist. Hier können wir für ein gegebenes y als $x = F^{-1}(y)$ irgendeinen x-Wert aus dem Intervall wählen, in dem die Verteilungsfunktion $F(x)$ konstant ist, und für jeden derartigen Wert von x erhalten wir dann $F[F^{-1}(y)] = y$. Insbesondere können wir für $x = F^{-1}(y)$ den kleinsten x-Wert wählen, für den diese Gleichung erfüllt ist.

Wir bezeichnen mit $F_1(y)$ die Verteilungsfunktion der Zufallsvariablen Y und erhalten

$$F_1(y) = P\big(F(X) < y\big) = \begin{cases} 0 & \text{für } y \leqq 0, \\ P\big(X < F^{-1}(y)\big) = F\big(F^{-1}(y)\big) = y & \text{für } 0 < y < 1, \\ 1 & \text{für } y \geqq 1. \end{cases}$$

Aus diesen Formeln folgt

$$F_1'(y) = f_1(y) = \begin{cases} 1 & \text{für} \quad 0 \leqq y \leqq 1, \\ 0 & \text{sonst.} \end{cases}$$

5.7. Die Normalverteilung

In Beispielen haben wir schon vielfach normal verteilte Zufallsvariable betrachtet. Nun wollen wir die allgemeine Form der Normalverteilung untersuchen.

Definition 5.7.1. Wir sagen, daß die Zufallsvariable X *normal verteilt* ist, wenn ihre Dichte $f(x)$ durch

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad (\sigma > 0) \tag{5.7.1}$$

gegeben ist.

Zuerst wollen wir nachprüfen, ob die durch (5.7.1) gegebene Funktion eine Dichte ist. Zu diesem Zweck transformieren wir X folgendermaßen:

$$Y = \frac{X - m}{\sigma}. \tag{5.7.2}$$

Wir erhalten

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}. \tag{5.7.3}$$

Da die durch (5.7.3) dargestellte Funktion eine Dichte ist (siehe 2.9), erhalten wir also

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = 1.$$

Die charakteristische Funktion $\varphi(t)$ der Zufallsvariablen Y wurde schon in Beispiel 4.2.2 berechnet:

$$\varphi(t) = e^{-\frac{t^2}{2}}.$$

Unter Benutzung der Gleichungen (4.2.14), (4.2.15) und (5.7.2) erhalten wir für die charakteristische Funktion der Zufallsvariablen X den Ausdruck

$$\varphi_1(t) = \exp\left(itm - \frac{\sigma^2 t^2}{2}\right). \tag{5.7.4}$$

Aus (5.7.4) und (4.2.4) erhalten wir für die Momente

$$m_1 = m, \quad m_2 = \sigma^2 + m^2, \quad \mu_2 = \sigma^2. \tag{5.7.5}$$

Wie man aus den Gleichungen (5.7.5) sieht, lassen sich die Konstanten m und σ, die in der Formel (5.7.1) vorkommen, leicht deuten: m ist der Mittelwert der Zufallsvariablen X und σ ihre Standardabweichung. Die Form der Dichtekurve einer normalen Zufallsvariablen hängt vom Wert von σ ab. Diese Kurven wollen wir im folgenden kurz *Normalkurven* nennen. Die Abhängigkeit von σ zeigt Abb. 5.7.1: Die drei Normalkurven haben denselben Mittelwert $m = 0$, aber verschiedene Standardabweichungen, nämlich $\sigma = 1$, $\sigma = 0{,}5$ und $\sigma = 0{,}25$.

Die Normalverteilung mit dem Mittelwert m und der Standardabweichung σ werden wir mit $N(m; \sigma)$ bezeichnen.

Wegen der Symmetrie der Normalkurve bezüglich des Mittelwertes m verschwinden alle ungeraden zentralen Momente:

$$\mu_{2k+1} = 0 \quad \text{für jedes } k. \tag{5.7.6}$$

Man zeigt leicht, daß

$$\mu_{2k} = 1 \cdot 3 \cdots (2k - 1)\sigma^{2k} \tag{5.7.7}$$

ist. Für $\sigma = 1$ ist die Formel (4.2.13) ein Spezialfall der Formel (5.7.7).

Für die Normalverteilung sind sehr genaue Tafeln vorhanden, die man bei Rechnungen benutzt. Gewöhnlich interessiert uns die Wahrscheinlichkeit dafür, daß die normale Zufallsvariable X einen Wert annimmt, der sich vom Mittelwert $m = E(X)$ absolut genommen um mehr als $\lambda\sigma$ $(\lambda > 0)$ unterscheidet, d. h. um

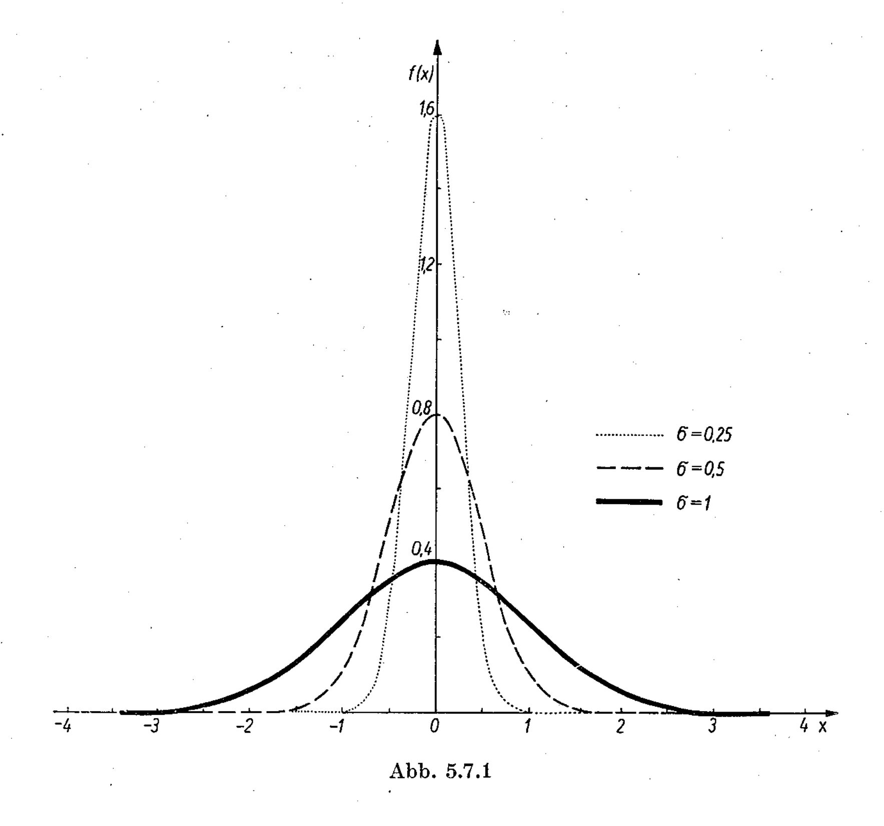


Abb. 5.7.1

mehr als ein gegebenes Vielfaches der Standardabweichung. Diese Wahrscheinlichkeit, ausgedrückt als Funktion von λ, finden wir in den Tafeln für die Normalverteilung, in denen die Werte des Integrals

$$P(|X - m| > \lambda\sigma) = \frac{2}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-\frac{y^2}{2}} dy$$

tabelliert sind. In der Tat, es ist

$$P(|X - m| > \lambda\sigma) = P\left(\frac{|X - m|}{\sigma} > \lambda\right) = P(|Y| > \lambda), \quad Y = \frac{X - m}{\sigma}.$$

Wir können auch fragen, wie groß die Wahrscheinlichkeit dafür ist, daß X um ein gegebenes Vielfaches $\lambda\sigma$ der Standardabweichung größer ist als der Mittelwert, d. h. wie groß die Wahrscheinlichkeit $P(X > m + \lambda\sigma)$ ist. Wir erhalten

$$P(X > m + \lambda\sigma) = P(Y > \lambda) = \frac{1}{\sqrt{2\pi}} \int\limits_{\lambda}^{\infty} e^{-\frac{y^2}{2}}\, dy.$$

Beispiel 5.7.1. Die Zufallsvariable X habe die Verteilung $N(1;2)$. Gesucht ist die Wahrscheinlichkeit dafür, daß X absolut genommen größer als 3 ist.

Wir führen die normierte Zufallsvariable $Y = \frac{X-1}{2}$ ein und erhalten

$$\begin{aligned} P(|X| > 3) &= P(|2Y + 1| > 3) = P\left(\left|Y + \frac{1}{2}\right| > \frac{3}{2}\right) \\ &= P\left(Y + \frac{1}{2} < -\frac{3}{2}\right) + P\left(Y + \frac{1}{2} > \frac{3}{2}\right) \\ &= P(Y < -2) + P(Y > 1). \end{aligned}$$

Da Y eine normierte, nach $N(0;1)$ verteilte Zufallsvariable ist, gilt

$$P(Y < -2) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{-2} e^{-\frac{t^2}{2}}\, dt = \frac{1}{\sqrt{2\pi}} \int\limits_{2}^{\infty} e^{-\frac{t^2}{2}}\, dt \approx 0{,}023,$$

$$P(Y > 1) = \frac{1}{\sqrt{2\pi}} \int\limits_{1}^{\infty} e^{-\frac{t^2}{2}}\, dt \approx 0{,}159.$$

Die Werte dieser Integrale lesen wir aus den Tafeln für die Normalverteilung ab, und wir erhalten schließlich

$$P(|X| > 3) = 0{,}182.$$

Den Tafeln für die Normalverteilung entnehmen wir, daß für eine nach $N(m;\sigma)$ normal verteilte Zufallsvariable X die Beziehungen

$$\begin{aligned} P(|X - m| > \sigma) &\approx 0{,}3173, \\ P(|X - m| > 2\sigma) &\approx 0{,}0455, \\ P(|X - m| > 3\sigma) &\approx 0{,}0027 \end{aligned}$$

erfüllt sind. Aus diesen Gleichungen erkennt man, daß die Normalverteilung in der Umgebung ihres Mittelwertes stark konzentriert ist. Die Wahrscheinlichkeit, daß sich der Wert von X vom Mittelwert absolut genommen um mehr als 3σ unterscheidet, ist kleiner als 0,01. Wegen dieser Eigenschaft der Normalverteilung wenden viele Statistiker häufig die sogenannte *Drei-Sigma-Regel* an. Nach dieser Regel ist in einer beliebigen Verteilung die Wahrscheinlichkeit dafür, daß die absolute Differenz zwischen der Zufallsvariablen und ihrem Mittelwert größer als 3σ ist, sehr klein. Diese Regel sollte man aber nicht kritiklos anwenden, denn nach der Tschebyscheffschen Ungleichung weiß man tatsächlich nur, daß für eine beliebige Zufallsvariable X

$$P(|X - m| > 3\sigma) < \frac{1}{9}$$

ist. Die Drei-Sigma-Regel kann man also nur auf Verteilungen anwenden, die sich nicht bedeutend von der Normalverteilung unterscheiden. Es müssen also fast symmetrische Verteilungen sein, die ihr Maximum in der Nähe des Symmetriemittelpunktes haben.

Für die Normalverteilung gilt ebenfalls ein Additionssatz. In der Tat, X und Y seien unabhängige Zufallsvariable, und X sei nach $N(m_1;\sigma_1)$, Y nach $N(m_2;\sigma_2)$ verteilt. Die charakteristischen Funktionen dieser Verteilungen sind

$$\varphi_1(t) = \exp\left(m_1 it - \frac{1}{2}\,t^2\sigma_1^2\right),$$

$$\varphi_2(t) = \exp\left(m_2 it - \frac{1}{2}\,t^2\sigma_2^2\right).$$

Wegen der Unabhängigkeit von X und Y hat die Zufallsvariable $Z = X + Y$ die charakteristische Funktion

$$\varphi(t) = \exp\left[(m_1 + m_2)it - \frac{1}{2}\,(\sigma_1^2 + \sigma_2^2)t^2\right]. \tag{5.7.8}$$

Der Ausdruck (5.7.8) ist die charakteristische Funktion der Normalverteilung $N\left(m_1 + m_2;\, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$, was zu beweisen war.

Cramér [4] hat gezeigt (vgl. Aufgabe 5.14.12), daß auch die Umkehrung des Satzes gilt: *Sind X_1 und X_2 unabhängig und hat die Zufallsvariable $X = X_1 + X_2$ eine Normalverteilung, so ist auch jede der Variablen X_1 und X_2 normal verteilt.*

Der Satz von Cramér gilt für beliebig viele unabhängige Zufallsvariable.

Neben dem Satz von Cramér kennt man viele Sätze, die die Normalverteilung charakterisieren. Einen hiervon werden wir in 9.5.C angeben. Hier führen wir den *Satz von* Skitowitsch [1] an: *Es seien $X_1, X_2, \ldots, X_n$ unabhängig, und sie mögen*

sämtlich die gleiche nicht ausgeartete Verteilung aufweisen. Dann ist die Unabhängigkeit der durch die Formeln

$$L_1 = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n,$$

$$L_2 = b_1 X_1 + b_2 X_2 + \cdots + b_n X_n$$

definierten Zufallsvariablen L_1 *und* L_2 *mit*

$$\sum_{j=1}^{n} a_j b_j = 0, \quad \sum_{j=1}^{n} (a_j b_j)^2 \neq 0$$

notwendig und hinreichend dafür, daß die Zufallsvariablen $X_1, X_2, \ldots, X_n$ *normal verteilt sind.*

Für $n = 2$ wurde dieser Satz von Bernstein [4], Darmois [2] und Gnedenko [10] bewiesen. Hier seien noch die Arbeiten von Pólya [2] (vgl. Aufgabe 6.16.10), Kac [1], Marcinkiewicz [1] (vgl. Aufgabe 5.14.11), Linnik [1], Laha [3] sowie Basu und Laha [1] erwähnt.

Die Normalverteilung hat sehr große Bedeutung in Wahrscheinlichkeitsrechnung und Statistik. In der Natur und in der Technik trifft man häufig auf Verteilungen, die sich der Normalverteilung sehr nähern. Diese Erscheinung ist ein Untersuchungsgegenstand der Theorie der stochastischen Prozesse (siehe 8.9). Außerdem ist unter ziemlich allgemeinen Voraussetzungen die Normalverteilung die Grenzverteilung einer Summe von unabhängigen Zufallsvariablen, wenn die Anzahl der Summanden ins Unendliche wächst. Diese Fragen werden wir im nächsten Kapitel behandeln.

5.8. Die Gammaverteilung

In den Anwendungen werden wir häufig eine Verteilung benutzen, die mit der *Gammafunktion* eng zusammenhängt, welche für $p > 0$ durch

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x}\, dx \tag{5.8.1}$$

definiert ist.

Es ist bekannt, daß das Integral (5.8.1) gleichmäßig bezüglich p konvergiert. Also ist $\Gamma(p)$ eine stetige Funktion. Wenn wir (5.8.1) partiell integrieren, erhalten wir

$$\Gamma(p+1) = \int_0^\infty x^p e^{-x}\, dx = \left[-e^{-x} x^p\right]_0^\infty + p \int_0^\infty x^{p-1} e^{-x}\, dx,$$

also

$$\Gamma(p+1) = p\,\Gamma(p). \tag{5.8.2}$$

Insbesondere folgen aus der Gleichung (5.8.2) für $p = n$ (n eine natürliche Zahl) sukzessive die Gleichungen

$$\begin{aligned} \Gamma(n+1) &= n\Gamma(n), \\ \Gamma(n) &= (n-1)\Gamma(n-1), \\ &\cdots\cdots\cdots \\ \Gamma(2) &= 1\Gamma(1). \end{aligned} \tag{5.8.3}$$

Da

$$\Gamma(1) = \int_0^\infty e^{-x}\,dx = -\left[e^{-x}\right]_0^\infty = 1$$

ist, ergibt sich aus den Gleichungen (5.8.3)

$$\Gamma(n+1) = n!. \tag{5.8.4}$$

Wenn wir in (5.8.1) die Substitution $y = \frac{x}{a}$ $(a > 0)$ ausführen, erhalten wir

$$\frac{\Gamma(p)}{a^p} = \int_0^\infty y^{p-1} e^{-ay}\,dy. \tag{5.8.5}$$

Die Gleichung (5.8.5) ist auch dann richtig, wenn a eine komplexe Zahl ist, $a = b + ic$, wobei $b > 0$. Das brauchen wir hier aber nicht zu beweisen.

Es sei X eine Zufallsvariable mit der durch

$$f(x) = \begin{cases} 0 & \text{für} \quad x \leqq 0, \\ \dfrac{b^p}{\Gamma(p)}\, x^{p-1} e^{-bx} & \text{für} \quad x > 0 \end{cases} \tag{5.8.6}$$

gegebenen Dichte $f(x)$; dabei ist $b > 0$ und $p > 0$. Daß durch (5.8.6) eine Dichte gegeben wird, folgt unmittelbar aus (5.8.5), denn es ist

$$\int_{\infty}^{\infty} f(x)\,dx = \int_0^\infty \frac{b^p}{\Gamma(p)}\, x^{p-1} e^{-bx}\,dx = 1$$

und $f(x)$ eine nichtnegative Funktion.

Definition 5.8.1. Wir sagen, daß die Zufallsvariable X eine *Gammaverteilung* hat, wenn ihre Dichte durch (5.8.6) gegeben wird.

Abb. 5.8.1 stellt die Dichte einer Gammaverteilung für $p = 1$ und $b = 0{,}5$ dar (vgl. auch Abb. 9.4.1).

Wir wollen die charakteristische Funktion dieser Verteilung bestimmen:

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx = \frac{b^p}{\Gamma(p)} \int_0^{\infty} x^{p-1} e^{-(b-it)x}\, dx. \tag{5.8.7}$$

Da, wie schon erwähnt, die Gleichung (5.8.5) auch für $a = b + ic$, $b > 0$, richtig ist, ergibt sich also aus (5.8.7)

$$\varphi(t) = \frac{b^p}{\Gamma(p)} \cdot \frac{\Gamma(p)}{(b - it)^p} = \frac{1}{\left(1 - \frac{it}{b}\right)^p}. \tag{5.8.8}$$

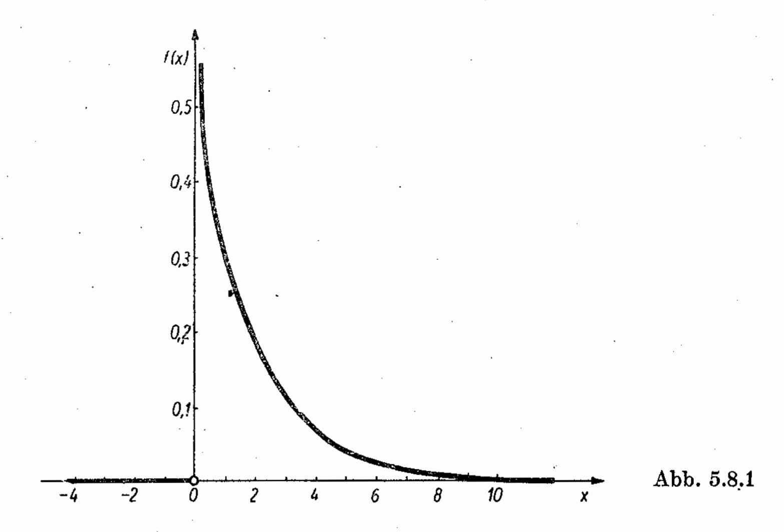


Abb. 5.8.1

Die Funktion $\varphi(t)$ kann man beliebig oft differenzieren. Ihre k-te Ableitung ist gleich

$$\varphi^{(k)}(t) = \frac{p(p+1)\cdots(p+k-1)}{b^k}\, i^k \frac{1}{\left(1 - \frac{it}{b}\right)^{p+k}} \qquad \text{für} \quad k = 1, 2, 3, \ldots$$

Aus Formel (4.2.4) ergibt sich

$$m_k = \frac{\varphi^{(k)}(0)}{i^k} = \frac{p(p+1)\cdots(p+k-1)}{b^k}. \tag{5.8.9}$$

Insbesondere erhalten wir

$$m_1 = \frac{p}{b}, \qquad m_2 = \frac{p(p+1)}{b^2}, \qquad \mu_2 = \frac{p}{b^2}. \tag{5.8.10}$$

Beispiel 5.8.1. Die Zufallsvariable X habe eine Gammaverteilung, deren Dichte durch

$$f(x) = \begin{cases} 0 & \text{für } x \leqq 0, \\ 2e^{-2x} & \text{für } x > 0 \end{cases}$$

gegeben wird. Wie der Leser nachprüfen kann, erhalten wir diese Verteilung, wenn wir in der Formel (5.8.6) $p = 1$ und $b = 2$ setzen. Wie groß ist die Wahrscheinlichkeit dafür, daß der Wert der Zufallsvariablen X nicht kleiner als 2 ist?

Wir erhalten

$$P(X \geqq 2) = 2\int_2^\infty e^{-2x}\,dx = -\left[e^{-2x}\right]_2^\infty = e^{-4} \approx 0{,}0183.$$

In komplizierteren Fällen kann man für die Gammaverteilung zur Berechnung der Wahrscheinlichkeit die Tafeln von PEARSON [1] heranziehen.

Die in Beispiel 5.8.1 behandelte Wahrscheinlichkeitsverteilung ist ein Spezialfall der Exponentialverteilung.

Definition 5.8.2. Wir sagen, daß die Zufallsvariable X eine *Exponentialverteilung* besitzt, wenn ihre Dichte $f(x)$ durch

$$f(x) = \begin{cases} 0 & \text{für } x \leqq 0, \\ \lambda e^{-\lambda x} & \text{für } x > 0, \end{cases} \tag{5.8.11}$$

wobei $\lambda > 0$, gegeben wird.

Wir zeigen noch, daß für Zufallsvariable mit Gammaverteilungen ein Additionssatz gilt.

Es seien X_1 und X_2 zwei unabhängige Zufallsvariable mit Gammaverteilungen mit den charakteristischen Funktionen

$$\varphi_k(t) = \frac{1}{\left(1 - \frac{it}{b}\right)^{p_k}} \qquad (k = 1, 2).$$

Wir betrachten die Veränderliche $X = X_1 + X_2$. Wegen der Unabhängigkeit der Veränderlichen X_1 und X_2 ist die charakteristische Funktion $\varphi(t)$ der Zufallsvariablen X gleich

$$\varphi(t) = \frac{1}{\left(1 - \frac{it}{b}\right)^{p_1}} \cdot \frac{1}{\left(1 - \frac{it}{b}\right)^{p_2}} = \frac{1}{\left(1 - \frac{it}{b}\right)^{p_1+p_2}}.$$

Daraus erkennt man, daß die Zufallsvariable X ebenfalls eine Gammaverteilung hat. Damit ist der Additionssatz bewiesen.

LAHA [1] und LUKACS [3] haben Sätze angegeben, die die Gammaverteilung charakterisieren. Wir geben hier den nachstehenden, besonders einfachen Satz von LUKACS an:

Die Zufallsvariablen X und Y mit nichtausgearteten Verteilungen seien unabhängig und mögen nur positive Werte annehmen. Die notwendige und hinreichende Bedingung dafür, daß diese Variablen eine Gammaverteilung mit dem gleichen Parameter b haben, ist die Unabhängigkeit der durch die Formeln

$$U = X + Y, \qquad V = \frac{X}{Y}$$

definierten Variablen U und V.

5.9. Die Betaverteilung

In den Anwendungen (vgl. 9.7) kommt auch eine Verteilung vor, die mit der Funktion

$$B(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,dx \qquad (p > 0,\ q > 0) \tag{5.9.1}$$

zusammenhängt.

Der Leser findet z. B. bei FICHTENHOLZ [1], S. 784, oder SMIRNOW [1], S. 223ff., den Beweis der Gleichung

$$B(p, q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}, \tag{5.9.2}$$

die die Funktion $B(p, q)$ mit der durch (5.8.1) definierten Funktion Γ verknüpft.

Definition 5.9.1. Wir sagen, daß die Zufallsvariable X eine *Betaverteilung* besitzt, wenn ihre Dichte $f(x)$ durch

$$f(x) = \begin{cases} \dfrac{1}{B(p, q)}\, x^{p-1}(1-x)^{q-1} & \text{für} \quad 0 < x < 1,\ p > 0,\ q > 0, \\ 0 & \text{für} \quad x \leqq 0 \quad \text{und} \quad x \geqq 1 \end{cases} \tag{5.9.3}$$

gegeben wird.

Daß die Funktion $f(x)$ eine Dichte darstellt, folgt daraus, daß sie nicht negativ ist, sowie aus den Formeln (5.9.1) und (5.9.2).

Es ist angebracht, die Momente der Betaverteilung unmittelbar aus der Formel

$$\begin{aligned} m_k &= \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} \int_0^1 x^{p+k-1}(1-x)^{q-1}\,dx \\ &= \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} \cdot B(p+k, q) = \frac{\Gamma(p+q)\,\Gamma(p+k)}{\Gamma(p)\,\Gamma(p+q+k)} \\ &= \frac{p(p+1)\cdots(p+k-1)}{(p+q)(p+q+1)\cdots(p+q+k-1)} \end{aligned} \tag{5.9.4}$$

zu berechnen. Insbesondere ist

$$m_1 = \frac{p}{p+q}, \tag{5.9.5}$$

$$m_2 = \frac{p(p+1)}{(p+q)(p+q+1)},$$

$$\mu_2 = \frac{pq}{(p+q)^2(p+q+1)}. \tag{5.9.6}$$

Abb. 5.9.1 stellt die Dichte der Betaverteilung für $p = q = 2$ dar.

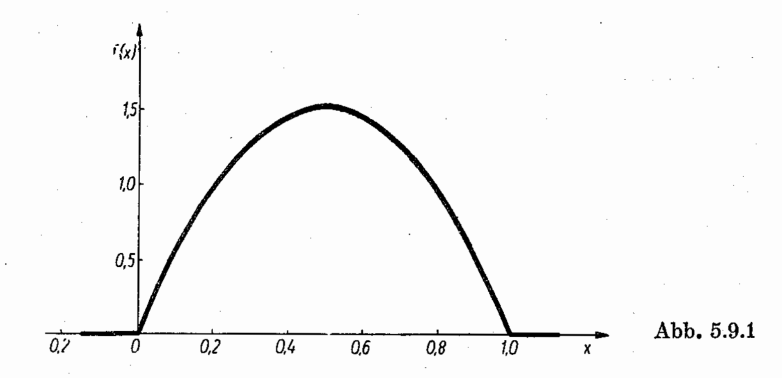


Abb. 5.9.1

Beispiel 5.9.1. Die Zufallsvariable Y habe eine Betaverteilung mit $p = q = 2$, die Dichte $f(y)$ dieser Verteilung hat also die Form

$$f(y) = \begin{cases} 0 & \text{für } y \leqq 0 \text{ und } y \geqq 1, \\ \dfrac{\Gamma(4)}{\Gamma(2)\Gamma(2)}\, y(1-y) = 6y(1-y) & \text{für } 0 < y < 1. \end{cases}$$

Wie groß ist die Wahrscheinlichkeit dafür, daß der Wert der Zufallsvariablen Y nicht größer als 0,2 ist?

Wir erhalten

$$P(Y \leqq 0{,}2) = 6 \int_0^{0{,}2} y(1-y)\,dy = 6\left[\frac{y^2}{2} - \frac{y^3}{3}\right]_0^{0{,}2} = 0{,}104.$$

Zur Berechnung der Wahrscheinlichkeiten für die Betaverteilung kann man die Tafeln von PEARSON [4] heranziehen.

5.10. Die Cauchy- und die Laplaceverteilung

A. Definition 5.10.1. Die Zufallsvariable X hat eine *Cauchyverteilung*, wenn ihre Dichte von der folgenden Form ist:

$$f(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x-\mu)^2} \quad (\lambda > 0). \tag{5.10.1}$$

Die Funktion $f(x)$ ist nicht negativ. Wenn wir die Substitution

$$y = \frac{x-\mu}{\lambda} \tag{5.10.2}$$

anwenden, erhalten wir

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{dy}{1+y^2} = \frac{1}{\pi}[\arctan y]_{-\infty}^{\infty} = 1.$$

Um die charakteristische Funktion der Zufallsvariablen X zu berechnen, suchen wir zuerst die charakteristische Funktion der durch die Formel (5.10.2) definierten Zufallsvariablen Y, die eine lineare Transformation der Zufallsvariablen X darstellt. Die Zufallsvariable Y hat die Dichte

$$f(y) = \frac{1}{\pi}\frac{1}{1+y^2} \tag{5.10.3}$$

und die charakteristische Funktion

$$\varphi(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{ity}\frac{1}{1+y^2}\,dy. \tag{5.10.4}$$

Um $\varphi(t)$ zu berechnen, betrachten wir zunächst die Dichte

$$f_1(y) = \frac{1}{2}e^{-|y|}. \tag{5.10.5}$$

Die charakteristische Funktion einer Zufallsvariablen mit der Dichte (5.10.5) hat die Form

$$\varphi_1(t) = \frac{1}{2}\int_{-\infty}^{\infty} e^{ity}e^{-|y|}dy = \frac{1}{2}\int_{-\infty}^{\infty}(\cos ty + i\sin ty)\,e^{-|y|}\,dy$$

$$= \int_0^{\infty}\cos ty\, e^{-y}\,dy.$$

Wenn wir partiell integrieren, erhalten wir

$$\int_0^\infty \cos ty e^{-y}\, dy = [-e^{-y} \cos ty]_0^\infty - t \int_0^\infty \sin ty e^{-y}\, dy = 1 - t \int_0^\infty \sin ty e^{-y}\, dy.$$

Ebenso ergibt sich

$$\int_0^\infty \sin ty e^{-y}\, dy = [-e^{-y} \sin ty]_0^\infty + t \int_0^\infty e^{-y} \cos ty\, dy = t \int_0^\infty e^{-y} \cos ty\, dy.$$

Daraus erhalten wir

$$\int_0^\infty e^{-y} \cos ty\, dy = 1 - t^2 \int_0^\infty e^{-y} \cos ty\, dy$$

und schließlich

$$\varphi_1(t) = \int_0^\infty e^{-y} \cos ty\, dy = \frac{1}{1+t^2}. \tag{5.10.6}$$

Die charakteristische Funktion (5.10.6) ist im Intervall $(-\infty, \infty)$ absolut integrierbar. Nach der Formel (4.5.6) entspricht ihr also eine Verteilung mit der Dichte

$$f_1(y) = \frac{1}{2\pi} \int_{-\infty}^\infty \frac{e^{-ity}}{1+t^2}\, dt. \tag{5.10.7}$$

Unter Beachtung der Formel (5.10.5) erhalten wir

$$e^{-|y|} = \frac{1}{\pi} \int_{-\infty}^\infty \frac{e^{-ity}}{1+t^2}\, dt.$$

Wenn wir unter dem Integralzeichen e^{-ity} durch e^{ity} ersetzen (diese Änderung hat keinen Einfluß auf den Wert des Integrals) und außerdem die Größen t und y miteinander vertauschen, erhalten wir

$$e^{-|t|} = \frac{1}{\pi} \int_{-\infty}^\infty \frac{e^{ity}}{1+y^2}\, dy. \tag{5.10.8}$$

Die rechte Seite dieser Formel ist identisch mit der rechten Seite von (5.10.4), und wir erhalten schließlich

$$\varphi(t) = e^{-|t|}. \tag{5.10.9}$$

Wenn wir beachten, daß X eine lineare Transformation der Zufallsvariablen Y ist, erhalten wir die charakteristische Funktion $\varphi_2(t)$ der Zufallsvariablen X in der Form

$$\varphi_2(t) = e^{i\mu t - \lambda |t|}. \tag{5.10.10}$$

Wie man leicht bestätigt, ist die Funktion $\varphi_2(t)$ im Punkt $t = 0$ nicht differenzierbar; es existiert also kein endliches Moment der Cauchyverteilung.

Für die Cauchyverteilung gilt ein Additionssatz. In der Tat, X_1 und X_2 seien zwei unabhängige Zufallsvariable mit Cauchyverteilungen, deren Dichten durch

$$g_1(x) = \frac{1}{\pi} \frac{\lambda_1}{\lambda_1^2 + (x - \mu_1)^2} \quad \text{bzw.} \quad g_2(x) = \frac{1}{\pi} \frac{\lambda_2}{\lambda_2^2 + (x - \mu_2)^2} \quad (\lambda_1, \lambda_2 > 0)$$

gegeben seien.

Die entsprechenden charakteristischen Funktionen sind

$$\varphi_1(t) = e^{i\mu_1 t - \lambda_1 |t|}, \qquad \varphi_2(t) = e^{i\mu_2 t - \lambda_2 |t|}.$$

Wir betrachten die Zufallsvariable $X = X_1 + X_2$. Ihre charakteristische Funktion wird durch

$$\varphi(t) = e^{i(\mu_1 + \mu_2)t - (\lambda_1 + \lambda_2)|t|} \tag{5.10.11}$$

gegeben. Diese Formel stellt ebenfalls die charakteristische Funktion einer Zufallsvariablen mit einer Cauchyverteilung dar.

B. Beim Beweis der Formel (5.10.9) betrachten wir eine Dichte, die durch die Formel (5.10.5) gegeben war. Die ihr entsprechende charakteristische Funktion wurde durch die Formel (5.10.6) gegeben. Die Zufallsvariable mit der Dichte (5.10.5) ist ein Spezialfall einer Zufallsvariablen mit einer Laplaceverteilung, und zwar hat X eine Laplaceverteilung, wenn $X = \lambda Y + \mu$ gilt, wobei Y die Dichte (5.10.5) hat. Nach (2.4.10) hat die Dichte von X die Form

$$f(x) = \frac{1}{2\lambda} \exp\left(-\frac{|x - \mu|}{\lambda}\right) \qquad (\lambda > 0). \tag{5.10.12}$$

Die charakteristische Funktion dieser Zufallsvariablen hat die Form

$$\varphi(t) = \frac{e^{\mu t i}}{1 + \lambda^2 t^2}. \tag{5.10.13}$$

Aus der Formel (5.10.13) folgt, daß eine Zufallsvariable, die eine Laplaceverteilung hat, endliche Momente beliebiger Ordnung besitzt.

5.11. Die n-dimensionale Normalverteilung

A. In Beispiel 3.6.1 betrachteten wir einen Spezialfall der zweidimensionalen Normalverteilung. Nun wollen wir kurz die allgemeine Form einer mehrdimensionalen Normalverteilung angeben und untersuchen.

Definition 5.11.1. Der Zufallsvektor (X, Y) hat eine *zweidimensionale* Normalverteilung, wenn seine Dichte $f(x, y)$ durch

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \tag{5.11.1}$$

$$\times \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2} - 2\frac{\varrho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}$$

gegeben wird, wobei m_1, m_2, σ_1, σ_2 und ϱ Konstanten sind, deren Bedeutung wir gleich besprechen wollen.

Man kann zeigen, daß die Gleichung

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$$

besteht.

Die Momente der ersten zwei Ordnungen sind

$$m_{10} = m_1, \quad m_{01} = m_2, \quad \mu_{20} = \sigma_1^2, \quad \mu_{02} = \sigma_2^2, \quad \mu_{11} = \varrho\sigma_1\sigma_2. \tag{5.11.2}$$

Die charakteristische Funktion dieser Verteilung hat die Form

$$\varphi(t, u) = \exp\left[i(m_1 t + m_2 u) - \frac{1}{2}(\sigma_1^2 t^2 + 2\varrho\sigma_1\sigma_2 t u + \sigma_2^2 u^2)\right]. \tag{5.11.3}$$

Wie wir sehen, hängt die Funktion $f(x, y)$ von fünf Konstanten ab. Übereinstimmend mit den Formeln (3.6.5), (3.6.9) und (3.6.26) sind m_1 und m_2 die Mittelwerte der Randverteilungen der Zufallsvariablen X und Y, die Konstanten σ_1 und σ_2 die Standardabweichungen dieser Randverteilungen, während ϱ den Korrelationskoeffizienten der Zufallsvariablen X und Y darstellt.

Wenn $\varrho^2 = 1$ ist, verliert der Ausdruck (5.11.1) seinen Sinn. Nach Satz 3.6.5 folgt jedoch aus der Gleichung $\varrho^2 = 1$, daß mit der Wahrscheinlichkeit Eins zwischen den Zufallsvariablen X und Y eine lineare Beziehung $Y = aX + b$ besteht. Es liegt dann eigentlich eine eindimensionale Verteilung vor, die *singuläre zweidimensionale Normalverteilung* genannt wird. In Zukunft wollen wir solche singulären Verteilungen nicht mehr betrachten.

Ist der Korrelationskoeffizient ϱ gleich Null, so nimmt die Formel (5.11.1) die Gestalt

$$
\begin{aligned}
f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\frac{(x-m_1)^2}{\sigma_1^2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\} \\
&= \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2}\right) \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(y-m_2)^2}{2\sigma_2^2}\right) \\
&= f_1(x)\, f_2(y)
\end{aligned}
\tag{5.11.4}
$$

an; dabei bedeuten $f_1(x)$ bzw. $f_2(y)$ die Dichten der Zufallsvariablen X bzw. Y.

Aus der Beziehung (5.11.4) folgt: Sind zwei Zufallsvariable mit einer zweidimensionalen Normalverteilung unkorreliert ($\varrho = 0$), so sind sie unabhängig. Wie man weiß, haben nicht alle zweidimensionalen Verteilungen diese Eigenschaft.

Die bedingte Dichte der Zufallsvariablen Y bei gegebenem $X = x$ hat laut Formel (2.7.6) die Gestalt

$$
f(y \mid x) = \frac{1}{\sigma_2\sqrt{1-\varrho^2}\,\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_2^2(1-\varrho^2)}\left[(y-m_2) - \frac{\varrho\sigma_2}{\sigma_1}(x-m_2)\right]^2\right\}.
\tag{5.11.5}
$$

Das ist die Dichte einer Normalverteilung, deren Mittelwert eine Funktion von x ist:

$$
m_2(x) = m_2 + \frac{\varrho\sigma_2}{\sigma_1}(x - m_1).
\tag{5.11.6}
$$

Die Formel (5.11.6) stellt die Gleichung der Regressionskurve erster Art von Y bezüglich X dar. Wie wir sehen, ist hier die Regressionskurve eine Gerade.

Ebenso erhalten wir für die Regressionskurve von X bezüglich Y die Gleichung

$$
m_1(y) = m_1 + \frac{\varrho\sigma_1}{\sigma_2}(y - m_2).
\tag{5.11.7}
$$

Einen Spezialfall der Regressionsgleichungen (5.11.6) und (5.11.7) betrachteten wir schon in Beispiel 3.7.2.

B. Wir wollen uns jetzt mit der n-dimensionalen Normalverteilung für $n \geqq 2$ beschäftigen.

Definition 5.11.2. Wir sagen, die Zufallsvariable $(X_1, X_2, \ldots, X_n)$ habe eine

n-dimensionale Normalverteilung, wenn ihre Dichtefunktion $f(x_1, x_2, \ldots, x_n)$ durch

$$f(x_1, x_2, \ldots, x_n) \tag{5.11.8}$$
$$= \frac{1}{(2\pi)^{n/2}\sqrt{|M|}} \exp\left[-\frac{1}{2|M|}\sum_{j,k=1}^{n} |M_{jk}|(x_j - m_j)(x_k - m_k)\right]$$

gegeben wird; dabei ist $|M| \neq 0$ die Determinante der Momentenmatrix M, deren Elemente λ_{jk} durch die Formel (3.6.30) gegeben werden, und $|M_{jk}|$ das algebraische Komplement des Elements λ_{jk} in $|M|$.

Wenn die Korrelationskoeffizienten aller Paare von Zufallsvariablen X_1, $X_2, \ldots, X_n$ gleich 0 sind, so ist $|M_{jk}| = 0$ für $j \neq k$, und die $X_1, X_2, \ldots, X_n$ sind voneinander unabhängig. Wir bemerken auch, daß wir es im Fall $|M| = 0$ nach Satz 3.6.6 mit einer Zufallsvariablen zu tun haben, deren Dimensionszahl mit der Wahrscheinlichkeit 1 kleiner als n ist.

Die charakteristische Funktion $\varphi(t_1, t_2, \ldots, t_n)$ einer Zufallsvariablen, deren Dichte durch (5.11.8) gegeben wird, hat die Form

$$\varphi(t_1, t_2, \ldots, t_n) = \exp\left(i\sum_{j=1}^{n} m_j t_j - \frac{1}{2}\sum_{j,k=1}^{n} \lambda_{jk} t_j t_k\right). \tag{5.11.9}$$

Einen Satz zur Charakterisierung der mehrdimensionalen Normalverteilung hat LAHA [2] angegeben.

Beispiel 5.11.1. Wir bezeichnen mit V die Geschwindigkeit eines Moleküls eines idealen Gases. Unseren Untersuchungen legen wir ein rechtwinkliges Koordinatensystem zugrunde. Es seien $\mathfrak{v}$ der Geschwindigkeitsvektor eines Moleküls und V_1, V_2, V_3 seine Projektionen auf die einzelnen Achsen. Die Lage des Vektors $\mathfrak{v}$ und seine Länge V sind vom Zufall abhängig; also ist (V_1, V_2, V_3) eine dreidimensionale Zufallsvariable. Um ihre Dichte $h(v_1, v_2, v_3)$ zu bestimmen, nahm MAXWELL die beiden folgenden Voraussetzungen an:

$$h(v_1, v_2, v_3) = f(v_1) f(v_2) f(v_3), \tag{*}$$

$$h(v_1, v_2, v_3) = h(v_1^2 + v_2^2 + v_3^2) = h(v^2). \tag{**}$$

In der Gleichung (*) bedeuten die Funktionen $f(v_1)$, $f(v_2)$, $f(v_3)$ die Dichten der Zufallsvariablen V_1, V_2, V_3. Die Voraussetzung (*) besagt also, daß diese Zufallsvariablen unabhängig sind und dieselben Verteilungen haben. Die Voraussetzung (**) besagt, daß die Dichte $h(v_1, v_2, v_3)$ eine Funktion der kinetischen Energie des Moleküls ist, die $\frac{m}{2}(v_1^2 + v_2^2 + v_3^2)$ beträgt, wobei m die Masse des Moleküls bedeutet. Aus diesen Gleichungen ergibt sich

$$\log h(v^2) = \sum_{i=1}^{3} \log f(v_i),$$

$$2\frac{h'(v^2)}{h(v^2)} \sum_{i=1}^{3} v_i\, dv_i = \sum_{i=1}^{3} \frac{f'(v_i)}{f(v_i)}\, dv_i.$$

Wenn wir die entsprechenden Ausdrücke bei dv_i miteinander vergleichen, erhalten wir

$$\frac{f'(v_1)}{v_1 f(v_1)} = \frac{f'(v_2)}{v_2 f(v_2)} = \frac{f'(v_3)}{v_3 f(v_3)} = k = \text{const.}$$

Die Auflösung dieser Differentialgleichungen ergibt

$$f(v_i) = C \exp\left(\frac{1}{2} k v_i^2\right) \qquad (i = 1, 2, 3).$$

Da $f(v_i)$ eine Dichte ist, muß $k < 0$ sein, wir können also $k = \frac{-1}{\sigma^2}$ schreiben. Unter Berücksichtigung der Formel (*) erhalten wir

$$h(v_1, v_2, v_3) = C^3 \exp\left(-\frac{v_1^2 + v_2^2 + v_3^2}{2\sigma^2}\right).$$

Die Zufallsvariable (V_1, V_2, V_3) hat also eine dreidimensionale Normalverteilung.

Da $h(v_1, v_2, v_3)$ eine Dichte ist, muß

$$C^3 = \frac{1}{(2\pi\sigma^2)^{3/2}}$$

sein. Die Größe σ ist für verschiedene Gase verschieden.

Zu der in diesem Beispiel behandelten Frage kehren wir im Beispiel 9.4.3 zurück.

C. Es sei $(X_1, X_2, \ldots, X_n)$ eine n-dimensionale Zufallsvariable mit Normalverteilung. Zur Vereinfachung der Rechnung setzen wir $m_j = 0$ $(j = 1, 2, \ldots, n)$. Wir betrachten die r-dimensionale Zufallsvariable $(Y_1, \ldots, Y_r)$, wobei

$$Y_l = \sum_{j=1}^{n} \alpha_{lj} X_j \qquad (l = 1, \ldots, r) \tag{5.11.10}$$

und $r \leqq n$ ist. Mit μ_{lh} bezeichnen wir die Kovarianz der Zufallsvariablen Y_l und Y_h $(l, h = 1, \ldots, r)$. Es ergibt sich

$$\begin{aligned} \mu_{lh} &= E(Y_l Y_h) = E\left(\sum_{j=1}^{n} \alpha_{lj} X_j \sum_{k=1}^{n} \alpha_{hk} X_k\right) \\ &= \sum_{j,k=1}^{n} \alpha_{lj} \alpha_{hk} E(X_j X_k) = \sum_{j,k=1}^{n} \alpha_{lj} \alpha_{hk} \lambda_{jk}. \end{aligned} \tag{5.11.11}$$

Wenn wir mit $\psi(u_1, \ldots, u_r)$ die charakteristische Funktion der Zufallsvariablen $(Y_1, \ldots, Y_r)$ bezeichnen, erhalten wir

$$\begin{aligned} \psi(u_1, \ldots, u_r) &= E\left[\exp\left(i \sum_{l=1}^{r} u_l Y_l\right)\right] \\ &= E\left[\exp\left(i u_1 \sum_{j=1}^{n} \alpha_{1j} X_j + \cdots + i u_r \sum_{j=1}^{n} \alpha_{rj} X_j\right)\right] \\ &= E\left[\exp\left(i \sum_{j=1}^{n} v_j X_j\right)\right] \end{aligned}$$

mit

$$v_j = u_1\alpha_{1j} + \cdots + u_r\alpha_{rj} \qquad (j = 1, 2, \ldots, n).$$

Berücksichtigen wir (5.11.9), so folgt

$$\begin{aligned}\psi(u_1, \ldots, u_r) &= \exp\left(-\frac{1}{2}\sum_{j,k=1}^{n} \lambda_{jk} v_j v_k\right) \\ &= \exp\left(-\frac{1}{2}\sum_{j,k=1}^{n} \lambda_{jk} \sum_{l,h=1}^{r} u_l u_h \alpha_{lj}\alpha_{hk}\right) \\ &= \exp\left(-\frac{1}{2}\sum_{l,h=1}^{r} u_l u_h \sum_{j,k=1}^{n} \alpha_{lj}\alpha_{hk}\lambda_{jk}\right).\end{aligned}$$

Daraus und aus (5.11.11) finden wir

$$\psi(u_1, \ldots, u_r) = \exp\left(-\frac{1}{2}\sum_{l,h=1}^{r} \mu_{lh} u_l u_h\right). \tag{5.11.12}$$

Wenn wir (5.11.12) mit (5.11.9) vergleichen, sehen wir, daß $(Y_1, \ldots, Y_r)$ eine r-dimensionale Normalverteilung hat. Wir haben also den folgenden wichtigen Satz erhalten:

Satz 5.11.1. *Es sei $(X_1, X_2, \ldots, X_n)$ eine n-dimensionale Zufallsvariable mit Normalverteilung, und $Y_1, \ldots, Y_r$ $(r \leq n)$ seien lineare Funktionen der Zufallsvariablen X_j $(j = 1, 2, \ldots, n)$. Dann ist die Zufallsvariable $(Y_1, \ldots, Y_r)$ ebenfalls normal verteilt.*

Aus diesem Satz folgt insbesondere, daß jede Randverteilung eines normal verteilten Zufallsvektors normal ist.

5.12. Die Polynomialverteilung

A. Wir wollen das verallgemeinerte Bernoullische Versuchsschema betrachten:

Ein Versuch werde n-mal wiederholt. Als Resultat eines Versuches kann eines der einander paarweise ausschließenden Ereignisse A_j $(j = 1, \ldots, r + 1)$ mit der Wahrscheinlichkeit $p_j = P(A_j)$, wobei $p_1 + \cdots + p_r + p_{r+1} = 1$ ist, eintreten. Bei n Versuchen kann das Ereignis A_j genau k_j-mal auftreten $(k_j = 0, 1, \ldots, n)$. Wir betrachten die Zufallsvariable $(X_1, \ldots, X_r, X_{r+1})$. Dabei bedeutet $X_j = k_j$, daß das Ereignis A_j genau k_j-mal auftrat. Die Ergebnisse der n Versuche sind unabhängig. Wenn wir ähnlich vorgehen wie in Beispiel 3.1.3 bei der Herleitung der Formel für die Binomialverteilung, so erhalten wir für die Wahrscheinlichkeitsfunktion der Zufallsvariablen $(X_1, \ldots, X_r, X_{r+1})$ die Formel

$$P(X_1 = k_1, \ldots, X_r = k_r, X_{r+1} = k_{r+1}) = \frac{n!}{k_1! \cdots k_r!\, k_{r+1}!}\, p_1^{k_1} \cdots p_r^{k_r} p_{r+1}^{k_{r+1}}, \tag{5.12.1}$$

dabei ist

$$k_1 + \cdots + k_r + k_{r+1} = n.$$

Die Formel gibt also die Wahrscheinlichkeit des Produkts an, d. h. die Wahrscheinlichkeit dafür, daß das Ereignis A_1 genau k_1-mal, das Ereignis A_2 genau k_2-mal, ..., das Ereignis A_{r+1} genau k_{r+1}-mal auftritt.

Wir bemerken, daß die Zufallsvariablen $X_1, X_2, \ldots, X_r, X_{r+1}$ der linearen Beziehung

$$X_1 + X_2 + \cdots + X_r + X_{r+1} = n$$

genügen. Eine dieser Zufallsvariablen, etwa X_{r+1}, drücken wir durch die anderen aus:

$$X_{r+1} = n - X_1 - X_2 - \cdots - X_r.$$

Die Formel (5.12.1) läßt sich nun in der Gestalt

$$P(X_1 = k_1, \ldots, X_r = k_r) = \frac{n!}{k_1! \cdots k_r! (n - K)!} p_1^{k_1} \cdots p_r^{k_r} q^{n-K} \tag{5.12.1'}$$

mit

$$K = k_1 + k_2 + \cdots + k_r \quad \text{und} \quad q = 1 - p_1 - p_2 - \cdots - p_r$$

schreiben.

Definition 5.12.1. Wir sagen, eine Zufallsvariable $(X_1, \ldots, X_r)$ mit der Wahrscheinlichkeitsfunktion (5.12.1) habe eine *Polynomialverteilung*.

B. Es seien

$$(Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_r^{(1)}) \quad \text{und} \quad (Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_r^{(2)})$$

zwei Zufallsvariable. Die Addition von mehrdimensionalen Zufallsvariablen wollen wir vektoriell verstehen, d. h., die Zufallsvariable $(X_1, X_2, \ldots, X_r)$ ist die Summe der Zufallsvariablen

$$(Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_r^{(1)}) \quad \text{und} \quad (Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_r^{(2)}),$$

wenn

$$X_j = Y_j^{(1)} + Y_j^{(2)} \qquad (j = 1, 2, \ldots, r);$$

dafür schreiben wir

$$(X_1, X_2, \ldots, X_r) = (Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_r^{(1)}) + (Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_r^{(2)}).$$

Es seien $(Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_r^{(m)})$ unabhängige Zufallsvektoren mit derselben

Verteilung, die höchstens eine von Null verschiedene Komponente haben, wobei für $m = 1, 2, \ldots, n$ und für $j = 1, 2, \ldots, r$

$$P(Y_j^{(m)} = 1) = p_j, \quad P(Y_j^{(m)} = 0) = 1 - p_j, \tag{5.12.2}$$

$$P(Y_1^{(m)} = 0, \ldots, Y_r^{(m)} = 0) = q = 1 - p_1 - \cdots - p_r$$

gelten soll. Wie man leicht einsieht, erfüllt eine Zufallsvariable $(X_1, X_2, \ldots, X_r)$ mit einer Polynomialverteilung die Gleichung

$$(X_1, X_2, \ldots, X_r) = \sum_{m=1}^{n} (Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_r^{(m)}).$$

Aus (4.6.1) erhalten wir, daß die charakteristische Funktion $\varphi_m(t_1, t_2, \ldots, t_r)$ der Zufallsvariablen $(Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_r^{(m)})$ für $m = 1, 2, \ldots, n$ die Form

$$\varphi_m(t_1, t_2, \ldots, t_r) = \sum_{j=1}^{r} p_j e^{it_j} + q$$

hat. Daraus erhalten wir für die charakteristische Funktion $\varphi(t_1, t_2, \ldots, t_r)$ der Zufallsvariablen $(X_1, X_2, \ldots, X_r)$ die Formel

$$\varphi(t_1, t_2, \ldots, t_r) = \prod_{m=1}^{n} \varphi_m(t_1, t_2, \ldots, t_r) = \left(\sum_{j=1}^{r} p_j e^{it_j} + q\right)^n. \tag{5.12.3}$$

Daraus und aus den zu (4.6.8') analogen Formeln folgt für $j = 1, 2, \ldots, r$

$$E(X_j) = np_j, \quad \lambda_{jj} = D^2(X_j) = np_j(1 - p_j) \tag{5.12.4}$$

und für $j, k = 1, 2, \ldots, r$, wenn $j \neq k$ ist,

$$\lambda_{jk} = E[(X_j - np_j)(X_k - np_k)] = -np_j p_k.$$

5.13. Zusammengesetzte Verteilungen

A. In den Anwendungen kommen häufig Zufallsvariable vor, deren Verteilung von einem gewissen Parameter α abhängt, der seinerseits wieder eine Zufallsvariable mit einer bestimmten Verteilung ist. Wir sagen dann, daß die Zufallsvariable X eine *zusammengesetzte Verteilung* hat. Hier wollen wir uns näher mit zwei zusammengesetzten Verteilungen befassen, nämlich mit der zusammengesetzten Binomialverteilung und der zusammengesetzten Poissonschen Verteilung.

Die Zufallsvariablen X_k $(k = 1, 2, \ldots)$ mögen unabhängig sein und sämtlich dieselbe Null-Eins-Verteilung haben, die durch die Wahrscheinlichkeiten $P(X_k = 1) = p$ und $P(X_k = 0) = 1 - p$ gegeben wird. Wir betrachten die

Zufallsvariable $X = X_1 + X_2 + \cdots + X_N$. Für festes N hat X die Binomialverteilung

$$P(X = s) = \binom{N}{s} p^s (1-p)^{N-s} \qquad (s = 0, 1, \ldots, N). \tag{5.13.1}$$

Es sei nun N eine von X_k $(k = 1, 2, \ldots)$ unabhängige Zufallsvariable, die nach dem Poissonschen Gesetz

$$P(N = n) = \frac{\lambda^n}{n!} e^{-\lambda} \qquad (n = 0, 1, 2, 3, \ldots) \tag{5.13.2}$$

verteilt ist. Wie wir sehen, spielt hier N die Rolle des erwähnten Parameters α. Wir betrachten die zweidimensionale Zufallsvariable (X, N) und erhalten (vgl. 2.7)

$$P(X = s,\ N = n) = P(X = s \mid N = n)\, P(N = n).$$

Uns interessiert die Wahrscheinlichkeit des Ereignisses $X = s$, wenn N alle möglichen Werte durchläuft. Mit anderen Worten, wir suchen die Randverteilung der Zufallsvariablen X. Diese Verteilung wird durch die Formel

$$P(X = s) = \sum_{n=0}^{\infty} P(X = s \mid N = n)\, P(N = n) \qquad (s = 0, 1, 2, \ldots)$$

gegeben. Bei Berücksichtigung der Formeln (5.13.1) und (5.13.2) erhalten wir, da $\binom{n}{s} = 0$ für $n < s$ ist,

$$\begin{aligned} P(X = s) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{s} p^s (1-p)^{n-s} \\ &= \frac{e^{-\lambda} p^s}{s!} \sum_{n=s}^{\infty} \frac{\lambda^n (1-p)^{n-s}}{(n-s)!} = \frac{e^{-\lambda} p^s \lambda^s}{s!} \sum_{n=s}^{\infty} \frac{[\lambda(1-p)]^{n-s}}{(n-s)!} \\ &= \frac{e^{-\lambda} p^s \lambda^s}{s!} e^{\lambda(1-p)} = \frac{e^{-\lambda p} (\lambda p)^s}{s!}. \end{aligned} \tag{5.13.3}$$

Das ist eine Poissonsche Verteilung mit dem Mittelwert $p\lambda$. Diese Verteilung nennen wir *zusammengesetzte Binomialverteilung*.

Beispiel 5.13.1. Die Wahrscheinlichkeit dafür, daß ein neugeborenes Kind ein Junge ist, beträgt 0,517 (vgl. Beispiel 1.1.3). Die Anzahl X der Geburten eines Jungen in einer Familie, in der N Kinder (N konstant) geboren wurden, ist eine binomial verteilte Zufallsvariable:

$$P(X = s) = \binom{N}{s} 0{,}517^s \cdot 0{,}483^{N-s} \qquad (s = 0, 1, \ldots, N).$$

Wir können aber auch nach der Wahrscheinlichkeit der Gleichung $X = s$ fragen, wenn N alle möglichen Werte annimmt, also nach der Wahrscheinlichkeit dafür, daß in einer (bezüglich der Zahl der Geburten) beliebigen Familie s Jungen geboren wurden. Hier ist N eine Zufallsvariable, deren Verteilung man für die Bevölkerung eines Landes empirisch bestimmen kann. Dazu braucht man nur den Anteil der kinderlosen Ehen, der Familien mit einem Kind, der Familien mit zwei Kindern usw. zu berechnen. Wenn wir die Verteilung von N kennen, können wir, indem wir ebenso wie bei der Herleitung der Formel (5.13.3) vorgehen, die Wahrscheinlichkeit $P(X = s)$ für jedes s berechnen.

B. Wir wollen uns jetzt mit der zusammengesetzten Poissonschen Verteilung beschäftigen.

Die Zufallsvariable X habe eine durch die Formel

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (k = 0, 1, 2, \ldots) \tag{5.13.4}$$

gegebene Poissonsche Verteilung, und λ sei eine stetige Zufallsvariable mit der Dichte

$$f(\lambda) = \begin{cases} \dfrac{a^v}{\Gamma(v)} \lambda^{v-1} e^{-a\lambda} & \text{für } \lambda > 0, \\ 0 & \text{für } \lambda \leqq 0, \end{cases} \tag{5.13.5}$$

wobei $v > 0$ und $a > 0$ ist.

Wir betrachten den Zufallsvektor (X, λ). Hier kommt eine bis jetzt noch nicht behandelte zweidimensionale Verteilung vor, in der die eine Zufallsvariable diskret und die andere stetig ist. Für jedes $h > 0$ und $\lambda_1 > 0$ erhalten wir

$$\begin{aligned} &P(X = k, \lambda_1 \leqq \lambda \leqq \lambda_1 + h) \\ &= P(X = k \mid \lambda_1 \leqq \lambda \leqq \lambda_1 + h)\, P(\lambda_1 \leqq \lambda \leqq \lambda_1 + h). \end{aligned}$$

Wir dividieren beide Seiten dieser Gleichung durch h und gehen für $h \to 0$ zur Grenze über. Das ist erlaubt, da $f(\lambda) > 0$ für $\lambda > 0$ ist. Wenn wir die Formeln (2.3.8) und (5.13.4) berücksichtigen, erhalten wir

$$\lim_{h \to 0} \frac{1}{h} P(X = k,\ \lambda_1 \leqq \lambda \leqq \lambda_1 + h) = \frac{\lambda_1^k}{k!} e^{-\lambda_1} \frac{a^v}{\Gamma(v)} \lambda_1^{v-1} e^{-a\lambda_1}. \tag{5.13.6}$$

Durch den Ausdruck (5.13.6) wird die Verteilung der zweidimensionalen Zufallsvariablen (X, λ) bestimmt. Die Randverteilung der Zufallsvariablen X erhalten wir aus der Formel (wir schreiben wieder λ statt λ_1)

$$P(X = k) = \int_0^\infty \frac{\lambda^k}{k!} e^{-\lambda} \frac{a^v}{\Gamma(v)} \lambda^{v-1} e^{-a\lambda}\, d\lambda.$$

Bei Beachtung der Formel (5.8.5) folgt daraus

$$P(X = k) = \frac{a^v}{\Gamma(v)} \int_0^\infty \frac{\lambda^{k+v-1} e^{-(a+1)\lambda}}{k!} \, d\lambda$$

$$= \frac{a^v}{\Gamma(v)} \cdot \frac{1}{k!} \frac{\Gamma(k+v)}{(a+1)^{k+v}} = \left(\frac{a}{1+a}\right)^v \frac{v(v+1)\cdots(v+k-1)}{(1+a)^k k!}. \tag{5.13.7}$$

Zur Vereinfachung der Schreibweise verallgemeinern wir die Bedeutung des Symbols $\binom{n}{r}$, das bis jetzt nur für natürliche n einen Sinn hatte. Für jedes x und jedes natürliche r definieren wir

$$\binom{x}{r} = \frac{x(x-1)\cdots(x-r+1)}{r!} = \frac{x_r}{r!}.$$

Wir führen weiter die Bezeichnungen $p = \frac{1}{1+a}$, $q = 1 - p = \frac{a}{1+a}$ ein. Nach Vorausssetzung ist $a > 0$, also gelten die Ungleichungen $0 < p < 1$ und $0 < q < 1$. Wenn wir diese Bezeichnungen benutzen, können wir (5.13.7) in der Form

$$P(X = k) = (-1)^k \binom{-v}{k} p^k q^v \qquad (k = 0, 1, 2, \ldots) \tag{5.13.8}$$

schreiben.

Definition 5.13.1. Wir sagen, daß die Zufallsvariable X eine *negative Binomialverteilung* hat, wenn ihre Wahrscheinlichkeitsfunktion durch die Formel (5.13.8) gegeben wird.

Wir berechnen die charakteristische Funktion dieser Verteilung:

$$\varphi(t) = \sum_{k=0}^\infty P(X = k)\, e^{itk} = q^v \sum_{k=0}^\infty (-1)^k \binom{-v}{k} p^k e^{itk}.$$

Wenn wir für $(1 - p)^{-v}$ die Reihenentwicklung

$$(1 - p)^{-v} = \sum_{k=0}^\infty (-1)^k \binom{-v}{k} p^k, \qquad |p| < 1,$$

benutzen, so erhalten wir

$$\varphi(t) = q^v (1 - p e^{it})^{-v}. \tag{5.13.9}$$

Daraus folgt

$$m_1 = \frac{\varphi'(0)}{i} = v\,\frac{p}{q},$$

$$m_2 = \frac{\varphi''(0)}{i^2} = \left(\frac{p}{q}\right)^2 v(v+1) + \frac{p}{q}\,v,$$

$$\mu_2 = m_2 - m_1^2 = v\,\frac{p}{q}\left(1 + \frac{p}{q}\right). \tag{5.13.10}$$

Für ein gewöhnliches Moment r-ter Ordnung ergibt sich

$$m_r = \sum_{l=0}^{r-1} \frac{(-1)^{r-l}\binom{r-1}{l}(-v)_{r-l}\,p^{r-l}}{q^{r-l}} \qquad (r = 1, 2, \ldots). \tag{5.13.11}$$

Greenwood und Yule [2] haben die zusammengesetzte Poissonsche Verteilung, die durch die Formel (5.13.8) definiert ist, in einer Reihe von Beispielen angewandt. Wir bringen hier eines dieser Beispiele:

Beispiel 5.13.2. Man hat im Laufe von drei Monaten die Anzahl der Arbeitsunfälle bei 414 Arbeitern beobachtet. Die Angaben befinden sich in Tabelle 5.13.1. Das Symbol k in der ersten Spalte bedeutet die Anzahl der Unfälle, die ein und demselben Arbeiter während der Dauer der Beobachtung zustießen. In der zweiten Spalte sind die beobachteten Häufigkeiten der Arbeiter, die während der Beobachtungszeit k Unfälle zu verzeichnen hatten, eingetragen,

Tabelle 5.13.1.

k	beobachtete Häufigkeit	Wahrscheinlichkeit
0	0,715	0,722
1	0,179	0,167
2	0,063	0,063
3	0,019	0,027
4	0,010	0,012
5	0,010	0,005
6	0,002	
7	0,000	0,004 (für 6–8)
8	0,002	

während sich in der dritten Spalte die entsprechenden Wahrscheinlichkeiten befinden, die nach der Formel (5.13.8) berechnet wurden. Die in dieser Formel vorkommenden unbekannten Parameter v und p hat man auf folgende Weise gefunden: Man berechnete den Mittelwert und die Dispersion der beobachteten Verteilung und nahm an, daß diese gleich $E(X)$ bzw. μ_2 sind. Für $E(X)$ und μ_2 gelten die Formeln (5.13.10). Auf diese Weise erhielt man zwei Gleichungen, aus denen man die unbekannten Parameter bestimmen konnte.

Wie wir sehen, unterscheiden sich die beobachteten Häufigkeiten wenig von den berechneten Wahrscheinlichkeiten. Diese Tatsache kann man folgendermaßen erklären:

Die Wahrscheinlichkeit dafür, daß ein Arbeiter während der Beobachtung k Unfällen unterliegt, wird durch die Poissonsche Verteilung bestimmt, in der der Parameter λ vorkommt. Auf die Größe dieses Parameters haben verschiedene von der Zeit abhängige Faktoren, wie der Stand der Unfallsicherheit, die atmosphärischen Bedingungen und andere, einen Einfluß. Man kann also sagen, daß λ eine Zufallsvariable ist. Unter der Annahme, daß λ eine Gammaverteilung hat, stellte man eine gute Übereinstimmung zwischen der beobachteten und der berechneten Verteilung fest.

C. Definition 5.13.2. Wenn die Zufallsvariable Y eine durch die Formel (5.13.8) gegebene Wahrscheinlichkeitsfunktion hat, wo v eine ganze Zahl ist, sagen wir, daß sie eine *Pascalverteilung* besitzt.

Diese Verteilung kann man auch auf Grund von völlig anderen Überlegungen erhalten, und zwar nicht als zusammengesetzte, sondern als einfache Verteilung.

Wir betrachten eine Folge von Versuchen. Als Ergebnis eines jeden Versuchs soll das Ereignis A oder das Ereignis $\bar{A}$ eintreten können. Die Ergebnisse jeder Anzahl von Versuchen sind unabhängig. Wir werden von einem „Erfolg" sprechen, wenn das Ereignis A eintritt, im entgegengesetzten Fall sprechen wir von einem „Mißerfolg". Es sei $P(A) = p$, also $P(\bar{A}) = q = 1 - p$.

Wir bezeichnen mit X_r die Anzahl der Erfolge, die nach dem $(r - 1)$-ten Mißerfolg eintraten und die dem r-ten Mißerfolg vorausgehen. So ist zum Beispiel X_1 die Anzahl der aufeinanderfolgenden Erfolge vor dem ersten Mißerfolg, X_2 die Anzahl der aufeinanderfolgenden Erfolge zwischen dem ersten und dem zweiten Mißerfolg. Wir betrachten die Zufallsvariable $X = X_1 + X_2 + \cdots + X_v$.

Das Ereignis $X = k$ ist das Produkt zweier Ereignisse: Das eine besteht darin, daß der $(k + v)$-te Versuch ein Mißerfolg ist, und das andere darin, daß unter den übrigen $k + v - 1$ Versuchen k Erfolge zu verzeichnen sind. Die Wahrscheinlichkeit des ersten Ereignisses ist gleich q und die des zweiten gleich $\binom{k+v-1}{k} p^k q^{v-1}$. Daraus erhalten wir

$$P(X = k) = \binom{k+v-1}{k} p^k q^v = \frac{v(v+1)\cdots(v+k-1)}{k!} p^k q^v$$

$$= (-1)^k \binom{-v}{k} p^k q^v \qquad (k = 0, 1, 2, \ldots). \tag{5.13.12}$$

Diese Formel ist identisch mit Formel (5.13.8).

D. Wir beweisen jetzt für die negative Binomialverteilung einen Satz, der das Analogon zu Satz 5.5.1 ist, der für die Binomialverteilung galt.

Satz 5.13.1. *Wenn für jedes v die Gleichung*

$$E(X) = v\frac{p}{q} = c$$

besteht, wobei c eine positive Konstante bedeutet, dann konvergiert die Wahrscheinlichkeitsfunktion der negativen Binomialverteilung für $v \to \infty$ gegen die entsprechende Funktion der Poissonverteilung.

Beweis. Aus Formel (5.13.8) erhalten wir

$$\lim_{v\to\infty} P(X = k) = \lim_{v\to\infty} \frac{v(v+1)\cdots(v+k-1)}{k!} p^k (1-p)^v$$
$$= \frac{c^k}{k!} \lim_{v\to\infty} \frac{v(v+1)\cdots(v+k-1)}{(v+c)^k} \left(1 - \frac{c}{v+c}\right)^v = \frac{e^{-c} c^k}{k!}. \tag{5.13.13}$$

Auf Grund dieser Formel kann man im Fall einer negativen Binomialverteilung die Tafeln für die Poissonverteilung benutzen.

E. Wir betrachten jetzt eine Zufallsvariable Y, die durch die Formel

$$Y = \sum_{k=1}^{N} X_k \tag{5.13.14}$$

gegeben wird; dabei sind X_k $(k = 1, 2, \ldots)$ und N Zufallsvariable, und N nimmt nur natürliche Zahlen als Werte an.

Satz 5.13.2. *Die Zufallsvariable N sei von den Veränderlichen $X_1, X_2, X_3, \ldots$ unabhängig und habe den Mittelwert $E(N)$. Wenn die Ungleichung*

$$\sum_{k=1}^{\infty} P(N \geqq k)\, E(|X_k|) < \infty \tag{5.13.15}$$

erfüllt ist, existiert der Mittelwert der Zufallsvariablen Y, und es ist

$$E(Y) = \sum_{k=1}^{\infty} P(N \geqq k)\, E(X_k). \tag{5.13.16}$$

Beweis. Die Formeln (5.13.14) und (3.6.22) ergeben

$$E(Y) = \sum_{n=1}^{\infty} P(N = n)\, E(Y \mid N = n) = \sum_{n=1}^{\infty} P(N = n) \sum_{k=1}^{n} E(X_k)$$
$$= \sum_{k=1}^{\infty} E(X_k) \sum_{n=k}^{\infty} P(N = n) = \sum_{k=1}^{\infty} E(X_k)\, P(N \geqq k).$$

Wenn wir (5.13.15) berücksichtigen, erhalten wir die Behauptung des obigen Satzes.

Setzen wir noch zusätzlich voraus, daß die Zufallsvariablen X_k dieselbe Verteilung haben, so ist die Bedingung (5.13.15) erfüllt, wenn der Mittelwert der

Zufallsvariablen X_k existiert. Wir wollen ihn mit $E(X)$ bezeichnen. Die Formel (5.13.16) hat dann die Gestalt

$$E(Y) = E(X) \sum_{k=1}^{\infty} P(N \geqq k) = E(X) \sum_{k=1}^{\infty} k P(N = k) = E(X)\, E(N). \tag{5.13.17}$$

Wie wir sehen, genügt der Mittelwert der zusammengesetzten Binomialverteilung der Gleichung (5.13.17).

5.14. Aufgaben und Ergänzungen

1. a) Man zeige, daß die Zufallsvariablen X_1 und X_2 unabhängig sind, wenn sie beide die Null-Eins-Verteilung aufweisen und nicht korreliert sind.
   b) Man überprüfe, ob alle Zweipunktvariablen diese Eigenschaft aufweisen.

2. Die Zufallsvariable X möge die Binomialverteilung (5.2.1) haben. Wir setzen

$$A_{nk} = \frac{P(X = k + 1)}{P(X = k)} \quad (k = 0, 1, \ldots, n - 1).$$

   Man beweise die folgenden Behauptungen:
   a) Die Ausdrücke $A_{nk} - 1$ und $(n + 1)p - 1 - k$ sind beide entweder gleich Null oder beide negativ oder beide positiv.
   b) $P(X = k)$ nimmt den maximalen Wert entweder in einem Punkt k_0 an, der der Ungleichung

$$(n + 1)p - 1 < k_0 < (n + 1)p$$

   genügt, wobei $(n + 1)p$ keine ganze Zahl ist, oder aber in den Punkten $(n + 1)p - 1$ und $(n + 1)p$, wenn $(n + 1)p$ eine ganze Zahl ist.
   c) Für $k > (n + 1)p$ gilt die Ungleichung

$$A_{nk} < \exp\left(- \frac{k + (n + 1)p}{n}\right).$$

3. Man beweise, daß die Ungleichung

$$\sum_{l=0}^{k} \binom{n_1}{l} \binom{n_2}{k - l} = \binom{n_1 + n_2}{k}$$

   gilt.
   Hinweis. Man verwende den Satz über die Addition Zufallsvariabler, die der Binomialverteilung genügen.

4. Man beweise die Beziehung

$$\sum_{m=k}^{n} \binom{n}{m} p^m (1 - p)^{n-m} = \frac{n!}{(k - 1)!\,(n - k)!} \int_0^p t^{k-1} (1 - t)^{n-k}\, dt.$$

Hinweis. Man integriere partiell und vergleiche die Formeln (10.3.7) und (10.3.8).

5. a) Man zeige, daß für beliebige $\lambda_1 > 0$ und $\lambda_2 > 0$ sowie für ganzzahlige nichtnegative k die Gleichung

$$\sum_{l=0}^{k} \frac{\lambda_1^l \lambda_2^{k-l}}{l!(k-l)!} = \frac{(\lambda_1 + \lambda_2)^k}{k!}$$

gilt.

b) Es seien X_1 und X_2 unabhängige Zufallsvariable mit Poissonscher Verteilung mit den Parametern λ_1 und λ_2. Ferner sei $Z = X_2 - X_1$. Man bestimme $P(Z = k)$ $(k = 0, \pm 1, \pm 2, \ldots)$ und zeige, daß für den Spezialfall $\lambda_1 = \lambda_2 = \lambda$

$$P(Z = k) = e^{-2\lambda} I_k(2\lambda)$$

gilt, wobei $I_k(2\lambda)$ die Besselsche Funktion mit rein imaginärem Argument ist.

6. Die unabhängige Variable Z möge eine verallgemeinerte Binomialverteilung haben. Man zeige (in den Bezeichnungen von 5.3): Sind die p_k als Funktionen von n so beschaffen, daß die Summe $\sum_{k=1}^{n} p_k = \lambda$ konstant ist und daß $a_n = \max(p_1, p_2, \ldots, p_n)$ für $n \to \infty$ gegen Null strebt, so gilt

$$\lim_{n\to\infty} P(Z = r) = e^{-\lambda} \frac{\lambda^r}{r!} \qquad (r = 0, 1, 2, \ldots).$$

(Dieses Ergebnis stammt von von Mises [5], [6]. Wegen eines stärkeren Ergebnisses siehe Hodges und LeCam [1].)

7. a) Es sei $F(x)$ die Verteilungsfunktion einer Zufallsvariablen X mit Null-Eins-Verteilung. Man bestimme die Verteilung der Zufallsvariablen $Y = F(X)$.

b) Man löse die gleiche Aufgabe für eine Zufallsvariable X mit Binomialverteilung.

c) Man löse die gleiche Aufgabe für X mit Poissonscher Verteilung.

8. Man sagt, die Zufallsvariable X habe eine *logarithmische Normalverteilung*, wenn ihre Dichte die Form

$$f(x) = \begin{cases} 0 & \text{für } x \leqq c, \\ \dfrac{1}{(x-c)\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(\log(x-c)-m)^2}{2\sigma^2}\right) & \text{für } x > c \end{cases}$$

hat, wobei c eine gewisse Konstante bedeutet. Man bestimme $E(X)$ und $D^2(X)$.

9. a) Man zeige, daß für jedes $x > 0$

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \left(\frac{1}{x} - \frac{1}{x^3}\right) < 1 - \Phi(x) < \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$$

gilt, wobei $\Phi(x)$ die Verteilungsfunktion der Normalverteilung $N(0;1)$ ist.

b) Man ermittle die entsprechende Ungleichung für $x < 0$.

10. Die Zufallsvariablen X_i $(i = 1, 2, 3)$ seien unabhängig und mögen die gleiche Verteilung $N(0;1)$ haben. Man bestimme die Verteilungsfunktion der Zufallsvariablen

$$Y = \max_{1 \leqq i \leqq 3} |X_i| .$$

11. Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien unabhängig und mögen die gleiche Verteilungsfunktion $F(x)$ mit endlichen Momenten beliebig hoher Ordnung haben. Die Größen $a_1, a_2, \ldots, a_n$ und $b_1, b_2, \ldots, b_n$ mögen den Beziehungen

$$\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j, \qquad \sum_{j=1}^{n} a_j^2 = \sum_{j=1}^{n} b_j^2$$

genügen, wobei die Folge $a_1, a_2, \ldots, a_n$ keine Permutation der Folge $b_1, b_2, \ldots, b_n$ ist. Man zeige, daß die Verteilungsfunktion $F(x)$ dann und nur dann normal ist, wenn die Zufallsvariablen L_1 und L_2,

$$L_1 = \sum_{j=1}^{n} a_j X_j, \qquad L_2 = \sum_{j=1}^{n} b_j X_j,$$

die gleiche Verteilung haben (Marcinkiewicz [1]).

12. Man beweise den Satz von Cramér [4] (siehe 5.7).
Hinweis. $F_1(x_1)$, $F_2(x_2)$, $\varphi_1(t)$ und $\varphi_2(t)$ mögen die Verteilungsfunktionen und die charakteristischen Funktionen der Zufallsvariablen X_1 und X_2 bezeichnen. $\Phi\big((x - m)/\sigma\big)$ sei die Normalverteilung der Summe $X_1 + X_2$. Aus der Ungleichung

$$F_1(x_1)\, F_2(x_2) \leqq \Phi\left(\frac{x_1 + x_2 - m}{\sigma}\right)$$

erhalten wir für $x_1 < 0$

$$F_1(x_1) < A \exp\left(-x_1^2/2\sigma^2 + B|x_1|\right)$$

und für $x_1 > 0$

$$1 - F_1(x_1) < A' \exp\left(-x_1^2/2\sigma^2 + B' x_1\right).$$

Aus diesen Ungleichungen folgt die Existenz des Integrals (vgl. 6.5)

$$I = \int_{-\infty}^{\infty} e^{x^2/4\sigma^2}\, dF_1(x).$$

Hieraus erhalten wir, daß $\varphi_1(t)$ als Funktion des komplexen Arguments t eine ganze Funktion von höchstens zweiter Ordnung ist. Die gleiche Eigenschaft weist $\varphi_2(t)$ auf. Anschließend verwende man den Satz von Hadamard und die Voraussetzung, daß die Summe $X_1 + X_2$ normal verteilt ist.

13. (Verallgemeinerung des Satzes von Cramér nach Linnik [1] und Zinger und Linnik [1].) Es seien die $\varphi_j(t)$ $(j = 1, 2, \ldots, n)$ charakteristische Funktionen von nichtausgearteten Verteilungen. Man zeige: Gilt die Gleichung

$$\prod_{j=1}^{n} \big(\varphi_j(t)\big)^{\alpha_j} = \exp\left(imt - \frac{1}{2}\,\sigma^2 t^2\right)$$

für ein gewisses System nichtnegativer Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_n$, so ist jede Funktion $\varphi_j(t)$ charakteristische Funktion einer normal verteilten Zufallsvariablen.

14. Es seien $X_1, X_2, \ldots, X_n$ unabhängig, und sie mögen die gleiche Verteilung haben. Man zeige, daß die Summe $X_1 + X_2 + \cdots + X_n$ genau dann von der Funktion $g(X_1, X_2, \ldots, X_n)$ unabhängig ist, wenn $g(X_1 + a, X_2 + a, \ldots, X_n + a)$ und $g(X_1, X_2, \ldots, X_n)$ für jedes a die gleiche Verteilung aufweisen (LAHA [3]).

15. Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien unabhängig und mögen die gleiche nicht ausgeartete, durch die Verteilungsfunktion $F(x)$ definierte Verteilung haben. Ferner sei

$$U = X_1 + \cdots + X_n, \qquad V = \sum_{r=1}^{n} \sum_{s=1}^{n} a_{rs} X_r X_s,$$

$$B_1 = \sum_{r=1}^{n} a_{rr}, \qquad B_2 = \sum_{r=1}^{n} \sum_{s=1}^{n} a_{rs}.$$

Man beweise die folgenden Behauptungen:

a) Ist $B_1 \neq 0$ und $B_2 = 0$, so ist $F(x)$ dann und nur dann die Verteilungsfunktion einer Normalverteilung, wenn die Zufallsvariable V eine konstante Regression (erster Art) bezüglich der Zufallsvariablen U hat, d. h., wenn $E(V|u) = E(V)$ mit der Wahrscheinlichkeit 1 gilt.

b) Ist $E(V) = 0$, $B_1 \neq 0$ und $B_2 \neq 0$, so ist $F(x)$ dann und nur dann Verteilungsfunktion einer Gammaverteilung, wenn die Zufallsvariable V eine konstante Regression bezüglich der Variablen U hat (vgl. LUKACS [3]).

16. Es seien $X_1, X_2, \ldots, X_n$ unabhängig, und sie mögen die gleiche Gammaverteilung haben. Man zeige (PITMAN [1]), daß die Summe $X_1 + X_2 + \cdots + X_n$ von einer beliebigen Funktion $g(X_1, X_2, \ldots, X_n)$ unabhängig ist, die gegenüber einer Skalentransformation invariant ist (d. h. für die $g(X_1, X_2, \ldots, X_n) = g(aX_1, aX_2, \ldots, aX_n)$ für jedes $a > 0$ gilt).

17. $X_1, X_2, \ldots, X_n$ seien unabhängig und mögen eine Gammaverteilung mit $b = 1$ aufweisen. Man zeige, daß die Summe $X_1 + X_2 + \cdots + X_n$ dann und nur dann von der Funktion $g(X_1, X_2, \ldots, X_n)$ unabhängig ist, wenn $g(X_1, X_2, \ldots, X_n)$ gegenüber einer Skalentransformation invariant ist (LAHA [3]).

18. Die Zufallsvariablen X_1 und X_2 seien unabhängig und mögen die Verteilung $N(0; 1)$ haben.

a) Man zeige, daß $Y = X_1/X_2$ eine Cauchysche Verteilung hat.

b) Man zeige, daß diese Eigenschaft die Normalverteilung nicht charakterisiert.

c) Man bestimme die Klasse der Zufallsvariablen, für die die in a) angegebene Bedingung erfüllt ist (LAHA [6], MAULDON [1] und KOTLARSKI [1]).

19. Man gebe ein Beispiel für eine zweidimensionale Zufallsvariable (X, Y) an, die keine Normalverteilung hat und deren beide Randverteilungen normal sind.
(Wegen einer allgemeinen Diskussion der mehrdimensionalen Verteilungsfunktionen mit vorgegebenen Randverteilungen siehe FRÉCHET [3] und GUMBEL [3].)

20. Man gebe ein Beispiel zweier Zufallsvariabler X und Y an, die beide normal verteilt sind und deren Summe $X + Y$ keine Normalverteilung besitzt. Man gebe insbesondere ein Beispiel dafür, daß die Summe $X + Y$ eine diskrete Verteilung hat.

21. Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien unabhängig und mögen die gleiche Verteilung haben, wobei die Dichte $f(x_1, x_2, \ldots, x_n)$ nur eine Funktion der Quadratsumme $x_1^2 + x_2^2 + \cdots + x_n^2$ ist. Man zeige, daß dann $X_1, X_2, \ldots, X_n$ normal verteilt sind. Hinweis. Man vergleiche hierzu Beispiel 5.11.1.

22. a) Man sagt, eine Zufallsvariable X habe eine *Paretoverteilung*, wenn ihre Dichte $f(x)$ durch die Formel

$$f(x) = \begin{cases} 0 & \text{für } x \leqq x_0, \\ \dfrac{a}{x_0}\left(\dfrac{x_0}{x}\right)^{a+1} & \text{für } x > x_0 \end{cases}$$

gegeben ist, wobei x_0 und $a > 0$ gewisse Konstanten sind. Man bestimme den Erwartungswert $E(X)$ und stelle fest, ob dieser stets existiert. Des weiteren bestimme man die Mediane der Zufallsvariablen X.

b) Wir sagen, daß die Zufallsvariable X eine *logistische Verteilung* hat, wenn ihre Verteilungsfunktion durch die Formel

$$F(x) = [1 + e^{-(ax+b)}]^{-1}$$

gegeben ist, wobei $a > 0$ ist. Man bestimme die charakteristische Funktion und die Momente dieser Verteilungsfunktion. Ferner zeige man, daß ihre Dichte $f(x)$ der Gleichung

$$f(x) = aF(x)[1 - F(x)]$$

genügt.

(Wegen der Verallgemeinerung der logistischen Verteilung auf den Fall zweier Dimensionen siehe Gumbel [4].)

23. Im Satz 5.13.2 sei die Voraussetzung der Unabhängigkeit von N und $X_1, X_2, \ldots$ durch die Voraussetzung ersetzt, daß das Ereignis $N = n$ von den Zufallsvariablen X_k für $k > n$ unabhängig ist. Man zeige, daß der Satz auch dann noch gültig bleibt (Kolmogoroff und Prochorow [1]).

24. Wahrscheinlichkeitsverteilungen, deren Dichtefunktionen $y = f(x)$ der Differentialgleichung

$$y' = \frac{x + a}{b_0 + b_1 x + b_2 x^2}\, y$$

genügen, wobei a, b_0, b_1, b_2 Konstanten sind, nennt man *Pearsonsche Verteilungen* (Pearson [6]–[8]).

a) Man zeige, daß die Normalverteilung, die Betaverteilung, die Gammaverteilung und die Paretoverteilung Pearsonsche Verteilungen sind.

b) Man zeige, daß die Konstanten a, b_0, b_1, b_2 sich als Funktionen der Momente erster bis vierter Ordnung ausdrücken lassen (falls diese existieren).

25. Eine Zufallsvariable X heißt *unbeschränkt teilbar*, wenn sich X für jedes natürliche n als Summe $X = X_1 + X_2 + \cdots + X_n$ darstellen läßt, wobei die $X_k (k = 1, 2, \ldots, n)$ unabhängig sind und die gleiche, von n abhängige Verteilungsfunktion $F_n(x)$ haben. Die Ver-

teilungsfunktion einer unbeschränkt teilbaren Zufallsvariablen heißt *unbeschränkt teilbare Verteilungsfunktion.*

a) Man definiere die unbeschränkte Teilbarkeit in den Termini der charakteristischen Funktionen.

b) Man zeige, daß normal verteilte Zufallsvariable sowie Zufallsvariable mit Poissonscher, Gaußscher und Cauchyscher Verteilung unbeschränkt teilbar sind.

c) Ist eine Zufallsvariable mit Binomialverteilung unbeschränkt teilbar?

26. Man zeige, daß die Summe einer beliebigen endlichen Anzahl unbeschränkt teilbarer Zufallsvariabler ebenfalls unbeschränkt teilbar ist.

27. Man zeige, daß die charakteristische Funktion $\varphi(t)$ einer unbeschränkt teilbaren Zufallsvariablen an keiner Stelle verschwindet.

Hinweis. Die unbeschränkte Teilbarkeit zieht für jedes n die Gleichung $\varphi(t) = (\varphi_n(t))^n$ nach sich, wobei $\varphi_n(t)$ ebenfalls eine charakteristische Funktion ist. Die Funktionen $\varphi(t)$ und $\varphi_n(t)$ sind in einer gewissen Umgebung von $t = 0$ von Null verschieden, d. h. für $|t| \leqq a$ für eine gewisses $a > 0$. Anschließend zeige man, daß für jedes $\varepsilon > 0$ ein n derart existiert, daß für alle t ($|t| \leqq a$) die Ungleichung $1 - |\varphi_n(t)| < \varepsilon$ gilt. Sodann greife man auf die in den Aufgaben 4.8.5a und 4.8.7b angegebenen Ergebnisse zurück, um zu zeigen, daß sich für jedes $a > 0$ ein n so finden läßt, daß für alle t, für die $|t| \leqq a$ gilt, der Ausdruck $1 - |\varphi_n(t)|$ beliebig klein ist. Dieses Ergebnis erweitere man auf Intervalle $|t| \leqq ra$ bei beliebigem ganzzahligem r.

28. Es sei $\varphi(t)$ die charakteristische Funktion einer unbeschränkt teilbaren Verteilungsfunktion. Man zeige, daß dann $(\varphi(t))^c$ für jedes $c > 0$ eine charakteristische Funktion ist. (Eine ausführliche Darstellung der Theorie der unbeschränkt teilbaren Verteilungen findet der Leser in dem Buch von Lévy [3] sowie in dem Buch von Gnedenko und Kolmogorov [1].)

6. GRENZWERTSÄTZE

6.1. Einleitende Bemerkungen

In der bisherigen Darstellung trafen wir schon einige Male auf Grenzwertsätze für Folgen von Wahrscheinlichkeitsfunktionen. So war im Satz 5.4.1 die Rede von der Konvergenz einer Folge von Wahrscheinlichkeitsfunktionen der Pólyaschen Verteilung gegen die entsprechende Funktion die Binomialverteilung und in Satz 5.5.1 von der Konvergenz einer Folge von Binomialverteilungen gegen die entsprechende Poissonsche Verteilung. Diese Sätze sind Spezialfälle von Grenzwertsätzen, die eine große theoretische und praktische Bedeutung haben.

Wir unterscheiden lokale und globale Grenzwertsätze. In den lokalen Sätzen untersucht man entweder (wie zum Beispiel in den Sätzen 5.4.1 und 5.5.1) den Grenzwert einer Folge von Wahrscheinlichkeitsfunktionen diskreter Zufallsvariabler oder den Grenzwert einer Folge von Dichtefunktionen stetiger Zufallsvariabler. Dagegen untersucht man in den globalen Grenzwertsätzen den Grenzwert einer Folge von Verteilungsfunktionen. In diesem Kapitel werden wir uns fast ausschließlich mit Grenzverteilungen von Summen unabhängiger Zufallsvariabler befassen. Wir beweisen einige wichtige Sätze über die Konvergenz der Verteilungsfunktionen von normierten Summen unabhängiger Zufallsvariabler gegen die Normalverteilung. Es sind dies die Sätze von Moivre-Laplace (global), Lindeberg-Lévy, Ljapunoff und Lindeberg-Feller. Außerdem bringen wir einen lokalen Grenzwertsatz von Gnedenko, wonach eine Folge von Wahrscheinlichkeitsfunktionen einer Summe unabhängiger Zufallsvariabler, die nur Werte der Form $a + hk$ ($h > 0$, k ganzzahlig) annehmen, gegen die Dichtefunktion der Normalverteilung konvergiert. Dieser Satz enthält als Spezialfall den lokalen Satz von Moivre-Laplace. Außer den Sätzen über die Konvergenz gegen die Normalverteilung beweisen wir einige Sätze, sogenannte „Gesetze der großen Zahlen“, in denen die Grenzverteilung die Einpunktverteilung ist.

Die moderne Theorie der Grenzverteilungen von Summen unabhängiger Zufallsvariabler hat sich in den letzten 30 Jahren bedeutend entwickelt, hauptsächlich auf Grund der Arbeiten von Chintschin [2] bis [5], Gnedenko [2] bis [4], Kolmogoroff [1], [2], [5] und Lévy [1] bis [3]. Es wurde eine einheitliche allgemeine Theorie geschaffen, in der die von uns angegebenen Grenzwertsätze Spezialfälle von allgemeineren Sätzen darstellen. Diese Sätze geben Konvergenzbedingungen für eine Folge von Verteilungsfunktionen von Summen von Zufallsvariablen, die weit allgemeiner als die von uns betrachteten Summen sind, gegen eine Grenz-

verteilungsfunktion an und legen auch die Klasse der möglichen, als Grenzverteilungen auftretenden Verteilungsfunktionen fest.

Eine ins einzelne gehende Behandlung dieser Theorie findet der Leser im Buch von Lévy [3] und in der Monographie von Gnedenko und Kolmogorov [1] sowie Loève [4].

Äußerst interessant sind auch die Konvergenzprobleme für Folgen von Verteilungsfunktionen einer Summe abhängiger Zufallsvariabler. Die ersten Arbeiten in dieser Richtung stammen von Markoff [1] und [3]. Wichtige Ergebnisse auf diesem Gebiet erzielte Bernstein [3]. Wir kehren zu dieser Frage noch in 7.5 zurück.

6.2. Die stochastische Konvergenz

A. Wir wollen das folgende Beispiel betrachten:

Beispiel 6.2.1. Die Zufallsvariable Y_n möge die Werte

$$0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1$$

annehmen, und ihre Wahrscheinlichkeitsfunktion sei durch

$$P\left(Y_n = \frac{r}{n}\right) = \binom{n}{r} \frac{1}{2^n} \qquad (r = 0, 1, \ldots, n) \tag{6.2.1}$$

gegeben.

Wir betrachten die durch

$$X_n = Y_n - \frac{1}{2} \tag{6.2.2}$$

definierte Zufallsvariable X_n.

Die Zufallsvariable X_n kann also die Werte

$$-\frac{1}{2}, \quad \frac{2-n}{2n}, \quad \frac{4-n}{2n}, \quad \ldots, \quad \frac{n-4}{2n}, \quad \frac{n-2}{2n}, \quad \frac{1}{2}$$

annehmen. Die Wahrscheinlichkeitsfunktion von X_n wird durch die folgende Formel gegeben:

$$P\left(X_n = \frac{2r-n}{2n}\right) = \binom{n}{r} \frac{1}{2^n}.$$

Es sei $n = 2$. Die Zufallsvariable X_2 kann die Werte

$$-0{,}5, \quad 0, \quad 0{,}5$$

mit den Wahrscheinlichkeiten

$$\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$$

annehmen. Es sei ε eine positive Zahl, etwa $\varepsilon = 0{,}3$; dann ist

$$P(|X_2| > 0{,}3) = P\left(X_2 = -\frac{1}{2}\right) + P\left(X_2 = \frac{1}{2}\right) = 0{,}5.$$

Es sei ferner $n = 5$. Die Zufallsvariable X_5 kann die Werte

$$-0{,}5, \quad -0{,}3, \quad -0{,}1, \quad 0{,}1, \quad 0{,}3, \quad 0{,}5$$

mit den Wahrscheinlichkeiten

$$\frac{1}{32}, \ \frac{5}{32}, \ \frac{10}{32}, \ \frac{10}{32}, \ \frac{5}{32}, \ \frac{1}{32}$$

annehmen, also ist

$$P(|X_5| > 0{,}3) = 0{,}0625.$$

Es sei nun $n = 10$. Die Zufallsvariable X_{10} kann die Werte

$$-0{,}5, \quad -0{,}4, \quad -0{,}3, \quad -0{,}2, \quad -0{,}1, \quad 0{,}0, \quad 0{,}1, \quad 0{,}2, \quad 0{,}3, \quad 0{,}4, \quad 0{,}5$$

mit den Wahrscheinlichkeiten

$$\frac{1}{1024}, \ \frac{10}{1024}, \ \frac{45}{1024}, \ \frac{120}{1024}, \ \frac{210}{1024}, \ \frac{252}{1024}, \ \frac{210}{1024}, \ \frac{120}{1024}, \ \frac{45}{1024}, \ \frac{10}{1024}, \ \frac{1}{1024}$$

annehmen; also erhalten wir für $n = 10$

$$P(|X_{10}| > 0{,}3) \approx 0{,}02.$$

Wie wir sehen, ist für $n = 10$ die Wahrscheinlichkeit dafür, daß der absolute Betrag von X_n größer als die feste Zahl $\varepsilon = 0{,}3$ ist, sehr klein.

Aus einem Satz, den wir in 6.3 beweisen werden, folgt, daß für unser Beispiel

$$\lim_{n\to\infty} P(|X_n| > 0{,}3) = 0 \tag{6.2.3}$$

ist; dabei wird sich zeigen, daß für die Folge der Zufallsvariablen X_n, die durch die Formel (6.2.2) definiert sind, die Beziehung (6.2.3) für jedes $\varepsilon > 0$ gilt.

B. Ehe wir zu diesem Satz übergehen, wollen wir den Begriff der stochastischen Konvergenz definieren.

Definition 6.2.1. Die Folge der Zufallsvariablen X_n *konvergiert stochastisch*[1]) gegen Null, wenn für beliebige $\varepsilon > 0$ die Beziehung

$$\lim_{n\to\infty} P(|X_n| > \varepsilon) = 0 \tag{6.2.4}$$

erfüllt ist.

[1]) In der Theorie der reellen Funktion entspricht der stochastischen Konvergenz die „Konvergenz dem Maße nach".

An Stelle von „stochastischer Konvergenz" trifft man in der Literatur auch häufig die Bezeichnung „Konvergenz in (der) Wahrscheinlichkeit" an.

Wir weisen darauf hin, daß in dieser Definition keine Rede davon ist, daß die Zufallsvariable X_n im Sinne der klassischen Analysis gegen Null konvergiert. Wenn also die Folge der Zufallsvariablen X_n stochastisch gegen Null konvergiert, so folgt daraus noch nicht, daß man für jedes ε ein endliches $n = n_0$ finden kann derart, daß für alle $n > n_0$ die Beziehung $|X_n| \leqq \varepsilon$ erfüllt ist. Aus der Definition der stochastischen Konvergenz folgt nur, daß die Wahrscheinlichkeit des Ereignisses $|X_n| > \varepsilon$ für $n \to \infty$ gegen Null strebt.

Satz 6.2.1. *Es sei $F_n(x)$ $(n = 1, 2, \ldots)$ die Verteilungsfunktion einer Zufallsvariablen X_n. Die Folge $\{X_n\}$ konvergiert dann und nur dann stochastisch gegen Null, wenn die Folge $\{F_n(x)\}$ der Beziehung*

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & \textit{für} \quad x < 0, \\ 1 & \textit{für} \quad x > 0 \end{cases} \tag{6.2.5}$$

genügt.

Beweis. Die Folge $\{X_n\}$ möge stochastisch gegen Null konvergieren.

Aus (6.2.4) folgt, daß für ein beliebiges ε für $n \to \infty$

$$P(X_n < -\varepsilon) = F_n(-\varepsilon) \to 0, \tag{6.2.6}$$

$$P(X_n > \varepsilon) = 1 - P(X_n \leqq \varepsilon) = 1 - F_n(\varepsilon) - P(X_n = \varepsilon) \to 0 \tag{6.2.6'}$$

gilt.

Da wir zu jedem $\varepsilon > 0$ stets ein ε_1 so finden können, daß $0 < \varepsilon_1 < \varepsilon$ ist, folgt aus (6.2.4), daß $P(X_n = \varepsilon) \to 0$ für jedes $\varepsilon > 0$ gilt. Aus (6.2.6) erhalten wir somit

$$1 - F_n(\varepsilon) \to 0. \tag{6.2.7}$$

Wir ersetzen ε durch $-x$ in (6.2.6) und durch x in (6.2.7), wobei $x > 0$ ist, und finden (6.2.5).

Nun sei (6.2.5) erfüllt. Für ein beliebiges $\varepsilon > 0$ erhalten wir dann

$$\lim_{n\to\infty} P(X_n < -\varepsilon) = \lim_{n\to\infty} F_n(-\varepsilon) = 0,$$

$$\lim_{n\to\infty} P(X_n > \varepsilon) \leqq \lim_{n\to\infty} P(X_n \geqq \varepsilon) = \lim_{n\to\infty} \big(1 - F_n(\varepsilon)\big) = 0.$$

Aus diesen beiden Beziehungen folgt sofort (6.2.4), womit der Satz vollständig bewiesen ist.

Wie wir uns erinnern, hat eine Zufallsvariable X, die eine Einpunktverteilung mit $P(X = 0) = 1$ besitzt, eine durch die Formel

$$F(x) = \begin{cases} 0 & \text{für} \quad x \leqq 0, \\ 1 & \text{für} \quad x > 0 \end{cases} \tag{6.2.8}$$

definierte Verteilungsfunktion. Diese Verteilungsfunktion ist in jedem Punkt $x \neq 0$ stetig. Aus den Beziehungen (6.2.5) und (6.2.7) folgt, daß die Folge der Verteilungsfunktionen $F_n(x)$ in jedem Punkt $x \neq 0$ gegen die durch die Formel (6.2.8) definierte Verteilungsfunktion $F(x)$ (im gewöhnlichen Sinne) konvergiert. Abschließend können wir also zusammenfassen: Die Folge der Verteilungsfunktionen $F_n(x)$ einer stochastisch gegen Null konvergierenden Folge von Zufallsvariablen strebt in jedem Punkt $x \neq 0$ gegen die Verteilungsfunktion einer Zufallsvariablen mit einer Einpunktverteilung.

Da die Punkte $x \neq 0$ Stetigkeitsstellen dieser Verteilungsfunktionen sind, können wir das erhaltene Ergebnis auch so formulieren:

Eine Folge $\{X_n\}$ von Zufallsvariablen konvergiert dann und nur dann stochastisch gegen Null, wenn die Folge der Verteilungsfunktionen dieser Zufallsvariablen gegen die durch (6.2.8) *definierte Verteilungsfunktion $F(x)$ in jeder Stetigkeitsstelle der Funktion (im üblichen Sinne) konvergiert.*

Wir betonen nochmals, daß in einer Unstetigkeitsstelle der Verteilungsfunktion $F(x)$, d. h. im Punkt $x = 0$, die Folge $\{F_n(0)\}$ nicht gegen $F(0)$ zu konvergieren braucht.

Man kann auch die stochastische Konvergenz einer Folge von Zufallsvariablen X_n gegen eine feste Konstante $c \neq 0$ definieren. Wir verstehen darunter die stochastische Konvergenz der Folge $\{Y_n\} = \{X_n - c\}$ gegen Null.

Ähnlich kann man die stochastische Konvergenz einer Folge $\{X_n\}$ von Zufallsvariablen gegen eine Zufallsvariable X erklären. Darunter verstehen wir, daß die Folge $\{Z_n\} = \{X_n - X\}$ stochastisch gegen Null konvergiert.

6.3. Das Bernoullische Gesetz der großen Zahlen

Wir wollen jetzt den schon in 6.2 angekündigten Satz beweisen. Einen Spezialfall dieses Satzes stellt die Formel (6.2.3) dar. Wir bezeichnen mit $\{Y_n\}$ eine Folge von diskreten Zufallsvariablen, deren Wahrscheinlichkeitsfunktionen durch

$$P\left(Y_n = \frac{r}{n}\right) = \binom{n}{r} p^r (1-p)^{n-r} \tag{6.3.1}$$

definiert sind; dabei ist $0 < p < 1$, und r kann die Werte $0, 1, 2, \ldots, n$ annehmen. Wir führen weiter folgende Bezeichnung ein:

$$X_n = Y_n - p. \tag{6.3.2}$$

Satz 6.3.1. *Die durch die Bedingungen* (6.3.1) *und* (6.3.2) *bestimmte Folge von Zufallsvariablen X_n konvergiert stochastisch gegen Null, d. h., für jedes $\varepsilon > 0$ ist*

$$\lim_{n\to\infty} P(|X_n| > \varepsilon) = 0. \tag{6.3.3}$$

Beweis. Zum Beweis benutzen wir die Tschebyscheffsche Ungleichung. Auf Grund der Gleichungen (5.2.8) ist

$$E(X_n) = 0, \tag{6.3.4}$$

$$\sigma_n = \sqrt{D^2(X_n)} = \sqrt{\frac{p(1-p)}{n}}. \tag{6.3.5}$$

Wenn wir in die Tschebyscheffsche Ungleichung die Ausdrücke (6.3.4) und (6.3.5) einsetzen, erhalten wir

$$P\left(|X_n| > k\sqrt{\frac{p(1-p)}{n}}\right) \leqq \frac{1}{k^2}, \tag{6.3.6}$$

wobei k eine beliebige positive Zahl ist. Wir nehmen

$$k = \varepsilon\sqrt{\frac{n}{p(1-p)}}$$

und setzen diesen Wert in die Ungleichung (6.3.6) ein. Es ergibt sich dann

$$P(|X_n| > \varepsilon) \leqq \frac{p(1-p)}{n\varepsilon^2} < \frac{1}{n\varepsilon^2}. \tag{6.3.7}$$

Aus (6.3.7) folgt, daß für ein beliebiges festes $\varepsilon > 0$ die Beziehung (6.3.3) gilt, was zu beweisen war.

Den soeben bewiesenen Satz nennt man das *Bernoullische Gesetz der großen Zahlen.*

Dieses Gesetz der großen Zahlen kann man folgendermaßen deuten:

Wir führen n Versuche nach dem Bernoullischen Versuchsschema aus; dabei ist die Wahrscheinlichkeit, daß das Ereignis A eintrifft, gleich p. Das Gesetz der großen Zahlen besagt nun: Ist n genügend groß, dann ist die Wahrscheinlichkeit dafür, daß die Häufigkeit des Ereignisses A sich wenig von p unterscheidet, nahezu Eins.

In den nächsten Paragraphen werden wir andere Gesetze der großen Zahlen kennenlernen.

6.4. Die Konvergenz einer Folge von Verteilungsfunktionen

In 6.2 haben wir gezeigt: Konvergiert die Folge von Zufallsvariablen X_n stochastisch gegen Null, so strebt die Folge $\{F_n(x)\}$ der Verteilungsfunktionen dieser Zufallsvariablen gegen die Verteilungsfunktion (6.2.8) der Einpunktverteilung, und zwar in jeder Stetigkeitsstelle dieser Verteilungsfunktion.

Wir wollen uns jetzt allgemein mit Fragen der Konvergenz einer Folge von Verteilungsfunktionen befassen.

Definition 6.4.1. Die Folge der Verteilungsfunktionen $F_n(x)$ einer Folge von Zufallsvariablen X_n heißt *konvergent*, wenn eine Verteilungsfunktion $F(x)$ existiert, so daß für jede Stetigkeitsstelle von $F(x)$ die folgende Beziehung erfüllt ist:

$$\lim_{n\to\infty} F_n(x) = F(x). \tag{6.4.1}$$

Die Verteilungsfunktion $F(x)$ nennen wir *Grenzverteilungsfunktion*.

Wie dieser Definition zu entnehmen ist, verlangt man nicht, daß die Folge der Verteilungsfunktionen auch in den Unstetigkeitspunkten von $F(x)$ gegen $F(x)$ konvergiert.

Wir kehren nochmals zum Beispiel 6.2.1 zurück. Auf Grund von Satz 6.3.1 konvergiert die durch die Formel (6.2.2) definierte Folge von Zufallsvariablen X_n stochastisch gegen Null. Also konvergiert die Folge der Verteilungsfunktionen $F_n(x)$ dieser Zufallsvariablen gegen die durch die Formel (6.2.8) definierte Verteilungsfunktion $F(x)$. Diese Verteilungsfunktion ist im Punkt $x = 0$ unstetig. Wir überzeugen uns leicht, daß die Zahlenfolge $\{F_n(0)\}$ nicht gegen $F(0)$ konvergiert.

Wir betrachten die Teilfolge der Folge $\{F_n(0)\}$, die nur aus Gliedern mit ungeradem Index $n = 2k + 1$ besteht. Die Zufallsvariable X_{2k+1} kann die Werte

$$-\frac{1}{2}, \quad \frac{2-(2k+1)}{2(2k+1)}, \quad \frac{4-(2k+1)}{2(2k+1)}, \quad \ldots, \quad \frac{2k+1-4}{2(2k+1)}, \quad \frac{2k+1-2}{2(2k+1)}, \quad \frac{1}{2}$$

annehmen. Für jedes k ist die eine Hälfte dieser Werte kleiner und die andere größer als Null. Die Wahrscheinlichkeit, daß X_{2k+1} einen kleineren Wert als Null annimmt, beträgt 0,5. Damit ist für jedes k

$$P(X_{2k+1} < 0) = F_{2k+1}(0) = 0{,}5,$$

also, da $F(0) = 0$ ist,

$$\lim_{k\to\infty} F_{2k+1}(0) = 0{,}5 \neq F(0). \tag{6.4.2}$$

Daher kann die Folge $\{F_n(0)\}$ nicht gegen $F(0) = 0$ konvergieren. Es ist also

$$\lim_{n\to\infty} F_n(0) \neq F(0),$$

obwohl nach der Definition 6.4.1 die Folge der Verteilungsfunktionen des Beispiels 6.2.1 gegen die durch die Formel (6.2.8) definierte Verteilungsfunktion konvergiert.

Wir bemerken ausdrücklich, daß wir nur dann sagen, eine Folge von Verteilungsfunktionen konvergiere, wenn sie gegen eine Verteilungsfunktion konvergiert. Dieser Vorbehalt ist wesentlich, da es vorkommen kann, daß eine Folge von Verteilungsfunktionen gegen eine Funktion konvergiert, die keine Verteilungsfunktion ist.

Beispiel 6.4.1. Wir betrachten die Folge der Zufallsvariablen X_n, die eine durch

$$P(X_n = n) = 1 \qquad (n = 1, 2, 3, \ldots)$$

gegebene Einpunktverteilung haben. Die Verteilungsfunktion $F_n(x)$ der Zufallsvariablen X_n hat die Form

$$F_n(x) = \begin{cases} 0 & \text{für} \quad x \leqq n, \\ 1 & \text{für} \quad x > n. \end{cases}$$

Hier besteht die Beziehung

$$\lim_{n\to\infty} F_n(x) = 0 \qquad (-\infty < x < \infty).$$

Die Folge $\{F_n(x)\}$ konvergiert also nicht gegen eine Verteilungsfunktion.

Die Folge von Verteilungsfunktionen $F_n(x)$ konvergiere gegen die Verteilungsfunktion $F(x)$, a und b $(a < b)$ seien zwei beliebige Stetigkeitspunkte der Grenzverteilungsfunktion $F(x)$. Dann besteht die Beziehung

$$\lim_{n\to\infty} P(a \leqq X_n < b) = F(b) - F(a). \tag{6.4.3}$$

In der Tat, es ist

$$P(a \leqq X_n < b) = F_n(b) - F_n(a). \tag{6.4.4}$$

Aus der Voraussetzung, daß a und b Stetigkeitsstellen der Verteilungsfunktion $F(x)$ sind, folgen die Beziehungen

$$F_n(b) \to F(b), \qquad F_n(a) \to F(a). \tag{6.4.5}$$

Aus den Formeln (6.4.4) und (6.4.5) folgt (6.4.3).

Die Folge der Verteilungsfunktionen $F_n(x)$ konvergiere gegen die Verteilungsfunktion $F(x)$; $P_n(S)$ und $P(S)$ seien die entsprechenden Wahrscheinlichkeitsfunktionen. Man kann zeigen, daß die Beziehung

$$\lim_{n\to\infty} P_n(S) = P(S) \tag{6.4.6}$$

für jede Borelsche Menge S besteht, für die $P\left(\bar{S} \cap \overline{R - S}\right) = 0$ ist, wobei $\bar{A}$ hier die Menge der Berührungspunkte von A bedeutet und R die x-Achse ist.

Wir weisen darauf hin, daß sogar in dem Fall, daß die Grenzverteilungsfunktion überall stetig ist, Borelsche Mengen S existieren können, für die die Beziehung (6.4.6) nicht gilt. Dies ist stets der Fall, wenn $P\left(\bar{S} \cap \overline{R - S}\right) > 0$ ist. Das folgende interessante Beispiel verdanken wir ROBBINS [2].

Beispiel 6.4.2. Die Zufallsvariable X_n $(n = 1, 2, \ldots)$ möge die durch

$$f_n(x) = \begin{cases} \dfrac{2^n}{\varepsilon} & \text{für } \dfrac{i}{n} - \dfrac{\varepsilon}{n \cdot 2^n} < x < \dfrac{i}{n}, \\ 0 & \text{im übrigen Gebiet} \end{cases}$$

definierte Dichte $f_n(x)$ haben, wobei $i = 1, 2, \ldots, n$ und $0 < \varepsilon < 1$ ist. Die Verteilungsfunktion $F_n(x)$ der Zufallsvariablen X_n hat somit die Form

$$F_n(x) = \begin{cases} 1 & \text{für } x \leqq 0, \\ \dfrac{i-1}{n} & \text{für } \dfrac{i-1}{n} \leqq x \leqq \dfrac{i}{n} - \dfrac{\varepsilon}{n \cdot 2^n}, \\ \dfrac{i-1}{n} + 2^n \dfrac{\left(x - \dfrac{i}{n} + \dfrac{\varepsilon}{n \cdot 2^n}\right)}{\varepsilon} & \text{für } \dfrac{i}{n} - \dfrac{\varepsilon}{n \cdot 2^n} < x < \dfrac{i}{n}, \\ 1 & \text{für } x \geqq 1. \end{cases} \tag{6.4.7}$$

Für jedes x aus dem Intervall $I = [0, 1]$ haben wir also

$$0 \leqq x - F_n(x) \leqq \frac{1}{n}. \tag{6.4.8}$$

Berücksichtigen wir darüber hinaus die Werte, die $F_n(x)$ außerhalb des Intervalls I annimmt, so erhalten wir für jedes reelle x die Beziehung

$$\lim_{n\to\infty} F_n(x) = F(x) = \begin{cases} 0 & \text{für } x \leqq 0, \\ x & \text{für } 0 < x < 1, \\ 1 & \text{für } x \geqq 1. \end{cases} \tag{6.4.9}$$

Nun bezeichnen wir mit S_n die Menge, in der $f_n(x) > 0$ ist, und mit S_∞ die durch

$$S_\infty = \sum_{n=1}^{\infty} S_n$$

definierte Borelsche Menge. $P_n(S)$ und $P(S)$ mögen die Wahrscheinlichkeitsfunktionen bezeichnen, die den Verteilungsfunktionen $F_n(x)$ und $F(x)$ entsprechen. Für $n = 1, 2, \ldots$ gilt dann

$$P_n(S_n) = \int_{S_n} f_n(x)\, dx = 1.$$

Wegen $S_\infty \supset S_n$ erhalten wir $P_n(S_\infty) = 1$ und damit

$$\lim_{n\to\infty} P_n(S_\infty) = 1.$$

Andererseits haben wir

$$P(S_\infty) = \int_{S_\infty} dx \leqq \sum_{n=1}^{\infty} \left(n \frac{\varepsilon}{n \cdot 2^n}\right) = \varepsilon < 1 \tag{6.4.10}$$

und somit

$$\lim_{n\to\infty} P_n(S_\infty) \neq P(S_\infty),$$

obwohl die Grenzverteilungsfunktion $F(x)$, die durch (6.4.9) definiert wurde, überall stetig ist. Dies ist der Fall, weil

$$P\left(\overline{S_\infty} \cap \overline{R - S_\infty}\right) > 0$$

gilt. In der Tat[1]), wir haben

$$P\left(\overline{S_\infty} \cap \overline{R - S_\infty}\right) = P\left(\overline{S_\infty} \cap \overline{I - S_\infty}\right).$$

Ferner ist die Menge $I - S_\infty$ abgeschlossen und nirgends dicht in I und damit

$$I - S_\infty = \overline{I - S_\infty} \quad \text{sowie} \quad \overline{I - (I - S_\infty)} = \overline{S_\infty} = I.$$

Hieraus erhalten wir unter Berücksichtigung von (6.4.10)

$$P\left(\overline{S_\infty} \cap \overline{I - S_\infty}\right) = P(I \cap I - S_\infty) = P(I - S_\infty) \geqq 1 - \varepsilon.$$

B. Wir geben jetzt eine Verallgemeinerung der Definition 6.4.1 für den Fall von Zufallsvektoren.

Definition 6.4.2. Die Folge der Verteilungsfunktionen $\{F_n(x_1, x_2, \ldots, x_k)\}$ der Zufallsvektoren $\{X_{n1}, X_{n2}, \ldots, X_{nk}\}$ heißt *konvergent*, wenn eine Verteilungsfunktion $F(x_1, x_2, \ldots, x_k)$ derart existiert, daß an jeder Stetigkeitsstelle die Bedingung

$$\lim_{n\to\infty} F_n(x_1, x_2, \ldots, x_k) = F(x_1, x_2, \ldots, x_k) \tag{6.4.11}$$

erfüllt ist.

Ist (6.4.11) erfüllt und bezeichnen $P_n(S)$ und $P(S)$ die entsprechenden Wahrscheinlichkeitsfunktionen, so läßt sich leicht zeigen, daß für eine beliebige Borelsche Menge S in einem k-dimensionalen euklidischen Raum R^k, für die $P(\overline{S} \cap \overline{R^k - S}) = 0$ gilt, die Beziehung (6.4.6) erfüllt ist. Diese Beziehung gilt insbesondere für Stetigkeitsintervalle (vgl. 2.5.C).

C. Von großer Bedeutung für Anwendungen ist der nachstehende Satz (Sverdrup [1], Prochorow [3]), den wir ohne Beweis angeben:

Satz 6.4.1. *Es sei* $\{F_n(x_1, x_2, \ldots, x_k)\}$ $(n = 1, 2, \ldots)$ *eine Folge von Verteilungsfunktionen der Zufallsvariablen* $(X_{n1}, X_{n2}, \ldots, X_{nk})$, *und es seien* $F(x_1, x_2, \ldots, x_k)$ *bzw.* $P(S)$ *die Verteilungsfunktion bzw. die Wahrscheinlichkeitsfunktion der Zufallsvariablen* $(X_1, X_2, \ldots, X_k)$. *Die Beziehung* (6.4.11) *gilt dann und nur dann, wenn für eine beliebige Funktion* $g(x_1, x_2, \ldots, x_k)$, *die auf einer der Gleichung*

[1]) Informationen über die hier benutzten Begriffe finden wir z. B. bei Hausdorff [2].

$P(S) = 1$ *genügenden Menge S stetig ist, die Beziehung*

$$\lim_{n\to\infty} H_n(\alpha) = H(\alpha),$$

wobei $H_n(\alpha)$ bzw. $H(\alpha)$ die Verteilungsfunktionen von $g(X_{n1}, X_{n2}, \ldots, X_{nk})$ bzw. $g(X_1, X_2, \ldots, X_k)$ sind, in jeder Stetigkeitsstelle der Funktion $H(\alpha)$ erfüllt ist.

6.5. Das Stieltjessche Integral

In den weiteren Betrachtungen werden wir einen Satz benutzen, der von Lévy [3] und Cramér [1] bewiesen wurde und uns erlaubt, die Untersuchung des Konvergenzverhaltens der Folge $\{F_n(x)\}$ von Verteilungsfunktionen der Zufallsvariablen $\{X_n\}$ an ihren charakteristischen Funktionen $\varphi_n(t)$ durchzuführen. Dieser Satz hat große Bedeutung in der Wahrscheinlichkeitsrechnung.

Der Beweis des Satzes verlangt die Einführung des Begriffes des Stieltjesschen Integrals. Dabei wird sich auch zeigen, daß wir die bis jetzt gesondert behandelten diskreten und stetigen Verteilungen mit Hilfe des Stieltjesschen Integrals gemeinsam erfassen können.

Zunächst führen wir den Begriff einer Funktion von endlicher Variation ein. $F(x)$ sei eine im Intervall $[a, b]$ (dieses Intervall kann auch unendlich sein) definierte Funktion. Wir zerlegen das Intervall durch die Punkte

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

in Teilintervalle und bilden die Summe

$$T = \sum_{k=0}^{n-1} |F(x_{k+1}) - F(x_k)|.$$

Der Wert T hängt natürlich von der Zahl n und der Zerlegung des Intervalls $[a, b]$ in Teilintervalle ab.

Definition 6.5.1. Die obere Grenze der Menge aller möglichen Werte von T nennen wir die *totale Variation der Funktion $F(x)$ im Intervall* $[a, b]$.

Definition 6.5.2. Wenn die totale Variation der Funktion $F(x)$ im Intervall $[a, b]$ endlich ist, sagen wir, die Funktion $F(x)$ sei im Intervall $[a, b]$ von *endlicher Variation.*

Jede beschränkte und nichtabnehmende Funktion $F(x)$ ist von endlicher Variation. Der Ausdruck $F(x_{k+1}) - F(x_k)$ ist dann natürlich für beliebiges k und eine beliebige Zerlegung des Intervalls $[a, b]$ nichtnegativ. Es besteht also die Beziehung

$$T = \sum_{k=0}^{n-1} [F(x_{k+1}) - F(x_k)] = F(b) - F(a),$$

wobei nach Voraussetzung $F(b)$ und $F(a)$ endlich sind.

Daraus folgt, daß jede Verteilungsfunktion von endlicher Variation ist; die totale Variation der Verteilungsfunktion $F(x)$ beträgt nämlich

$$F(\infty) - F(-\infty) = 1.$$

Wir wollen jetzt den Begriff des Stieltjesschen Integrals erklären. In einem endlichen Intervall $[a, b]$ seien die Funktionen $g(x)$ und $F(x)$ gegeben. Dabei setzen wir voraus, daß die Funktion $F(x)$ von endlicher Variation und linksseitig stetig ist und keinen Sprung im Punkt b hat und daß ferner die Funktion $g(x)$ stetig ist.

Wir zerlegen das Intervall $[a, b]$ durch die Punkte

$$a = x_0 < x_1 < \cdots < x_n = b$$

in n Teilintervalle. Im k-ten Teilintervall $[x_k, x_{k+1})$ wählen wir einen beliebigen Punkt x_k' und bilden die Summe

$$S = \sum_{k=0}^{n-1} g(x_k')[F(x_{k+1}) - F(x_k)]. \tag{6.5.1}$$

Definition 6.5.3. Wenn die Summe S für $n \to \infty$ und $\max(x_{k+1} - x_k) \to 0$ gegen einen bestimmten Grenzwert I strebt, und zwar unabhängig von der Wahl der Punkte x_k' und der Zerlegung des Intervalls $[a, b]$, dann nennen wir diesen Grenzwert das *Stieltjessche Integral der Funktion* $g(x)$ *nach der Funktion* $F(x)$. Dieses Integral schreiben wir in der Form

$$I = \lim_{n\to\infty} S = \int_a^b g(x)\, dF(x). \tag{6.5.2}$$

Wie wir sehen, ist das Stieltjessche Integral eine Verallgemeinerung des Riemannschen Integrals: Für $F(x) = x$ stellt die Formel (6.5.2) ein Riemannsches Integral dar.

Im Fall eines unbegrenzten Integrationsintervalls definiert man das uneigentliche Stieltjessche Integral als Grenzwert einer Folge von eigentlichen Stieltjesschen Integralen. Wenn also der Grenzwert

$$\lim_{\substack{a\to-\infty\\ b\to\infty}} \int_a^b g(x)\, dF(x)$$

existiert, wobei a und b auf beliebige Weise gegen $-\infty$ bzw. ∞ streben, dann nennen wir diesen Grenzwert das *uneigentliche Stieltjessche Integral der Funktion* $g(x)$ *nach der Funktion* $F(x)$.

Wir führen ohne Beweis einige Eigenschaften des Stieltjesschen Integrals an:

1. Wenn c und l Konstante sind, dann ist

$$\int_a^b c g(x)\, d[l F(x)] = c l \int_a^b g(x)\, dF(x).$$

2. Existieren die Integrale auf der rechten Seite der beiden nachstehenden Gleichungen, so existieren auch die Integrale auf der linken Seite, und es ist

$$\int_a^b [g_1(x) + g_2(x)]\, dF(x) = \int_a^b g_1(x)\, dF(x) + \int_a^b g_2(x)\, dF(x),$$

$$\int_a^b g(x)\, d[F_1(x) + F_2(x)] = \int_a^b g(x)\, dF_1(x) + \int_a^b g(x)\, dF_2(x).$$

3. Wenn $a < c < b$ und die drei Integrale

$$\int_a^b g(x)\, dF(x), \quad \int_a^c g(x)\, dF(x), \quad \int_c^b g(x)\, dF(x)$$

sämtlich existieren, dann gilt die Gleichung

$$\int_a^b g(x)\, dF(x) = \int_a^c g(x)\, dF(x) + \int_c^b g(x)\, dF(x).$$

Die Eigenschaften 2 sind für Summen von beliebig vielen Funktionen $g_i(x)$ und $F_i(x)$ erfüllt, und Eigenschaft 3 gilt für beliebig, aber endlich viele Punkte $a < c_1 < c_2 < \cdots < c_n < b$.

Ist die Funktion $g(x)$ stetig und beschränkt auf der ganzen x-Achse, dann existiert sowohl jedes eigentliche als auch das uneigentliche Stieltjessche Integral. Das Stieltjessche Integral kann aber auch für unbeschränkte Funktionen $g(x)$ existieren.

Die Funktion $F(x)$ habe im Intervall $[a, b]$ die Ableitung $F'(x)$, und es existiere das Riemannsche Integral der Funktion $F'(x)$.

Wenden wir die Formel von Lagrange (Mittelwertsatz) an, so erhalten wir

$$\int_a^b g(x)\, dF(x) = \lim_{n\to\infty} \sum_{k=0}^{n-1} g(x_k')F'(x_k')(x_{k+1} - x_k) = \int_a^b g(x)F'(x)\, dx.$$

Unter diesen Bedingungen wird das Stieltjessche Integral auf das Riemannsche Integral der Funktion $g(x)F'(x)$ zurückgeführt. Insbesondere ist, wenn $F(x)$ die Verteilungsfunktion einer stetigen Zufallsvariablen X mit der Dichte $f(x)$ ist,

$$\int_a^b g(x)\, dF(x) = \int_a^b g(x)f(x)\, dx. \tag{6.5.3}$$

Es sei nun $F(x)$ die Verteilungsfunktion einer diskreten Zufallsvariablen mit den Sprungstellen $x_1', x_2', x_3', \ldots$ Die Verteilungsfunktion $F(x)$ ist dann durch die Formel

$$F(x) = \sum_{x_k' < x} [F(x_k' + 0) - F(x_k')]$$

gegeben. Wenn wir nun die Formeln (6.5.1) und (6.5.2) berücksichtigen, erhalten wir für $x_k \in [a, b)$

$$\int_a^b g(x)\, dF(x) = \sum_k g(x_k')\, p_k. \tag{6.5.4}$$

Aus (6.5.3) und (6.5.4) folgt: Existiert der Mittelwert der Zufallsvariablen $Y = g(X)$, dann wird er sowohl für eine stetige Zufallsvariable X als auch für eine diskrete Zufallsvariable durch die Formel

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, dF(x) \tag{6.5.5}$$

gegeben. Hierbei ist $F(x)$ die Verteilungsfunktion der Zufallsvariablen X.

Die bisher nebeneinander gebrauchten Formeln, die Integralformeln für stetige Zufallsvariable und die Reihenformeln für diskrete Zufallsvariable, kann man nun bei Benutzung des Stieltjesschen Integrals einheitlich in Gestalt einer einzigen Formel schreiben. So erhalten wir z. B., wenn wir in der Formel (6.5.5) $g(x) = x^r$ setzen, für ein Moment r-ter Ordnung den allgemeinen Ausdruck (vgl. den Anhang)

$$m_r = E(X^r) = \int_{-\infty}^{\infty} x^r\, dF(x).$$

Ebenso erhalten wir, wenn wir $g(x) = e^{itx}$ setzen, für die charakteristische Funktion

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x).$$

Existiert das Stieltjessche Integral (6.5.2) und ist die Funktion $g(x)$ an den Stellen a und b stetig, so läßt es sich nach der Formel

$$\int_a^b g(x)\, dF(x) = [g(x) F(x)]_a^b - \int_a^b F(x)\, dg(x)$$

durch partielle Integration berechnen.

Es lassen sich auch leicht die nachstehenden Formeln für die Verteilungsfunktionen einer Summe, einer Differenz, eines Produkts und eines Quotienten unabhängiger Zufallsvariabler X und Y ohne die Voraussetzung herleiten, daß der Zufallsvektor (X, Y) stetig ist. (Die Beweise dieser Formeln findet der Leser im Anhang.)

Es seien $F_1(x)$, $F_2(y)$ und $F(z)$ die Verteilungsfunktionen der Zufallsvariablen X, Y und Z. Dann gelten die folgenden Beziehungen:

I. Ist $Z = X + Y$, so gilt

$$F(z) = \int_{-\infty}^{\infty} F_2(z - x)\, dF_1(x) = \int_{-\infty}^{\infty} F_1(z - y)\, dF_2(y). \tag{6.5.6}$$

II. Ist $Z = X - Y$, so gilt

$$F(z) = \int_{-\infty}^{\infty} F_1(z + y)\, dF_2(y). \tag{6.5.7}$$

III. Ist $Z = XY$ mit $P(X = 0) = P(Y = 0) = 0$, so gilt

$$\begin{aligned} F(z) &= \int_{-\infty}^{0} [1 - F_2(z/x)]\, dF_1(x) + \int_{0}^{\infty} F_2(z/x)\, dF_1(x) \\ &= \int_{-\infty}^{0} [1 - F_1(z/y)]\, dF_2(y) + \int_{0}^{\infty} F_1(z/y)\, dF_2(y). \end{aligned} \tag{6.5.8}$$

IV. Ist $Z = \dfrac{X}{Y}$ mit $P(Y = 0) = 0$, so gilt

$$F(z) = \int_{-\infty}^{0} [1 - F_1(zy)]\, dF_2(y) + \int_{0}^{\infty} F_1(zy)\, dF_2(y). \tag{6.5.9}$$

6.6. Der Satz von Lévy und Cramér

Den obenerwähnten Satz von Lévy und Cramér stellen wir zuerst in Form von zwei Sätzen dar.

Satz 6.6.1a. *Konvergiert die Folge* $\{F_n(x)\}$ $(n = 1, 2, 3, \ldots)$ *von Verteilungsfunktionen gegen die Verteilungsfunktion* $F(x)$, *so konvergiert die entsprechende Folge* $\{\varphi_n(t)\}$ *der charakteristischen Funktionen an jeder Stelle* t $(-\infty < t < \infty)$ *gegen die Funktion* $\varphi(t)$, *wobei* $\varphi(t)$ *die charakteristische Funktion der Grenzverteilungsfunktion* $F(x)$ *ist; die Konvergenz ist dabei gleichmäßig in* t *in jedem endlichen Werteintervall von* t.

Beweis. Nach Definition der charakteristischen Funktion ist

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx}\, dF_n(x), \quad \varphi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x).$$

Sind $a < 0$ und $b > 0$ zwei Stetigkeitspunkte der Verteilungsfunktion $F(x)$, so erhalten wir

$$\varphi_n(t) = \int_{-\infty}^{a} e^{itx}\, dF_n(x) + \int_{a}^{b} e^{itx}\, dF_n(x) + \int_{b}^{\infty} e^{itx}\, dF_n(x) = I_{n1} + I_{n2} + I_{n3},$$

$$\varphi(t) = \int_{-\infty}^{a} e^{itx}\, dF(x) + \int_{a}^{b} e^{itx}\, dF(x) + \int_{b}^{\infty} e^{itx}\, dF(x) = I_1 + I_2 + I_3. \quad (6.6.1)$$

Wir integrieren die Differenz

$$I_{n2} - I_2 = \int_{a}^{b} e^{itx}\, dF_n(x) - \int_{a}^{b} e^{itx}\, dF(x)$$

partiell und erhalten

$$I_{n2} - I_2 = e^{itx}\{[F_n(x)]_a^b - [F(x)]_a^b\} - it \int_{a}^{b} [F_n(x) - F(x)] e^{itx}\, dx,$$

also

$$|I_{n2} - I_2| \leqq |F_n(b) - F(b)| + |F_n(a) - F(a)| + |t| \int_{a}^{b} |F_n(x) - F(x)|\, dx.$$

Es sei $\varepsilon > 0$ beliebig gewählt. Aus den Voraussetzungen des Satzes und daraus, daß a und b Stetigkeitsstellen der Verteilungsfunktion $F(x)$ sind, folgt für genügend große n

$$|F_n(b) - F(b)| < \frac{\varepsilon}{9}, \qquad |F_n(a) - F(a)| < \frac{\varepsilon}{9}.$$

Weiter folgt aus dem Lebesgueschen Satz vom Grenzübergang unter dem Integralzeichen (vgl. NATANSON [1]), wenn wir berücksichtigen, daß $|F_n(x) - F(x)|$ in jedem Intervall gleichmäßig beschränkt ist, die Gleichung

$$\lim_{n\to\infty} \int_{a}^{b} |F_n(x) - F(x)|\, dx = \int_{a}^{b} \lim_{n\to\infty} |F_n(x) - F(x)|\, dx.$$

Da die Funktion unter dem Integralzeichen auf der rechten Seite der letzten Gleichung gleich Null ist mit Ausnahme von (höchstens) abzählbar vielen Punkten, ist also das betrachtete Integral gleich Null. Es genüge t der Ungleichung $T_1 < t < T_2$, wobei T_1 und T_2 zwei beliebige, aber fest gewählte Zahlen sind, und es sei K das Maximum der beiden Zahlen $|T_1|$ und $|T_2|$, $K = \max(|T_1|, |T_2|)$. Für genügend große n und alle betrachteten t ist dann

$$|t| \int_{a}^{b} |F_n(x) - F(x)|\, dx \leqq K \int_{a}^{b} |F_n(x) - F(x)|\, dx < \frac{\varepsilon}{9}.$$

Also erhalten wir

$$|I_{n2} - I_2| < \frac{\varepsilon}{3}. \tag{6.6.2}$$

Weiter betrachten wir die Differenz

$$I_{n1} - I_1 = \int_{-\infty}^{a} e^{itx}\, dF_n(x) - \int_{-\infty}^{a} e^{itx}\, dF(x).$$

Es ist

$$|I_{n1} - I_1| \leqq \int_{-\infty}^{a} dF_n(x) + \int_{-\infty}^{a} dF(x) = F_n(a) + F(a).$$

Ist aber der absolute Wert von a hinreichend groß, so folgen für genügend große n aus der Stetigkeit von $F(x)$ im Punkt a und aus der Voraussetzung unseres Satzes die Beziehungen

$$F_n(a) < \frac{\varepsilon}{6}, \qquad F(a) < \frac{\varepsilon}{6};$$

also gilt für alle t und für genügend große n

$$|I_{n1} - I_1| < \frac{\varepsilon}{3}. \tag{6.6.3}$$

Ebenso folgt

$$|I_{n3} - I_3| < \frac{\varepsilon}{3}. \tag{6.6.4}$$

Aus den Formeln (6.6.1) bis (6.6.4) folgt die Behauptung des Satzes.

Satz 6.6.1b (Umkehrung). *Konvergiert die Folge von charakteristischen Funktionen $\varphi_n(t)$ an jeder Stelle t $(-\infty < t < \infty)$ gegen die Funktion $\varphi(t)$, die in einem gewissen Intervall $|t| < \tau$ stetig ist, dann konvergiert die Folge der entsprechenden Verteilungsfunktionen $F_n(x)$ gegen die Verteilungsfunktion $F(x)$, die die Funktion $\varphi(t)$ als charakteristische Funktion besitzt.*

Beweis. Zum Beweis benutzen wir den Hellyschen Satz (siehe Natanson [1]), der aussagt, daß jede Folge von Verteilungsfunktionen $F_n(x)$ eine konvergente Teilfolge $\{F_{n_k}(x)\}$ enthält; der Grenzwert dieser Teilfolge ist eine nichtfallende Funktion $F(x)$, die in den Unstetigkeitsstellen so abgeändert werden kann, daß sie linksseitig stetig wird. Aus dem Hellyschen Satz folgt aber noch nicht, daß $F(x)$ eine Verteilungsfunktion ist. Offenbar ist, da $F(x)$ der Grenzwert einer Folge von Verteilungsfunktionen ist, $0 \leqq F(x) \leqq 1$; man weiß aber nicht, ob $F(-\infty) = 0$ und $F(\infty) = 1$ ist. Wir zeigen, daß dies in der Tat der Fall ist.

Zu diesem Zweck nehmen wir an, das wäre nicht der Fall, so daß

$$a = F(\infty) - F(-\infty) < 1 \tag{6.6.5}$$

wäre. Da aber $\varphi_n(t) \to \varphi(t)$ und $\varphi_n(0) = 1$ ist, folgt $\varphi(0) = 1$. Weiter schließt man aus der Stetigkeit der Funktion $\varphi(t)$, daß in einer genügend kleinen Umgebung des Punktes $t = 0$ sich $\varphi(t)$ wenig von Eins unterscheidet, so daß für genügend kleines τ die Ungleichung

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} \varphi(t)\,dt\right| > 1 - \frac{\varepsilon}{2} > a + \frac{\varepsilon}{2} \tag{6.6.6}$$

erfüllt ist, wobei die positive Zahl ε so gewählt ist, daß $a + \varepsilon < 1$ ist. Da weiter die Teilfolge $\{F_{n_k}(x)\}$ gegen $F(x)$ konvergiert, folgt aus der Beziehung (6.6.5), daß man $b > \frac{4}{\varepsilon\tau}$ (wobei $-b$ und b Stetigkeitspunkte der Grenzverteilungsfunktion sind) und K so groß wählen kann, daß für $k > K$ die Beziehung

$$a_k = F_{\tau_k}(b) - F_{\tau_k}(-b) < a + \frac{\varepsilon}{4}$$

besteht.

Andererseits folgt aus der Beziehung (6.6.6) wegen $\varphi_n(t) \to \varphi(t)$, daß für genügend große k die Ungleichung

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} \varphi_{n_k}(t)\,dt\right| > a + \frac{\varepsilon}{2} \tag{6.6.7}$$

besteht. Wir zeigen, daß diese Ungleichung nicht erfüllt sein kann; denn es ist

$$\int_{-\tau}^{\tau} \varphi_{n_k}(t)\,dt = \int_{-\tau}^{\tau}\int_{-\infty}^{\infty} e^{itx}\,dF_{n_k}(x)\,dt = \int_{-\infty}^{\infty}\left[\int_{-\tau}^{\tau} e^{itx}\,dt\right] dF_{n_k}(x).$$

Wegen $|e^{itx}| = 1$ erhalten wir

$$\left|\int_{-\tau}^{\tau} e^{itx}\,dt\right| \leqq 2\tau. \tag{6.6.8}$$

Andererseits ist

$$\left|\int_{-\tau}^{\tau} e^{itx}\,dt\right| = \left|\left[\frac{e^{itx}}{ix}\right]_{-\tau}^{\tau}\right| = \frac{2}{|x|}\,|\sin\tau x| \leqq \frac{2}{|x|} < \frac{2}{b} \quad \text{für } |x| > b. \tag{6.6.9}$$

Wir teilen jetzt die ganze x-Achse in zwei Teile, nämlich in die Punktmenge $|x| \leqq b$ und in die außerhalb dieses Intervalls liegende Punktmenge. Es ergibt sich

$$\left| \int_{-\infty}^{\infty} \left[\int_{-\tau}^{\tau} e^{itx}\, dt \right] dF_{n_k}(x) \right| \leqq \left| \int_{|x| \leqq b} \left[\int_{-\tau}^{\tau} e^{itx}\, dt \right] dF_{n_k}(x) \right| + \left| \int_{|x| > b} \left[\int_{-\tau}^{\tau} e^{itx}\, dt \right] dF_{n_k}(x) \right|.$$

Wenn wir für $|x| \leqq b$ die Ungleichung (6.6.8) und für $|x| > b$ die Ungleichung (6.6.9) benutzen, erhalten wir

$$\frac{1}{2\tau} \left| \int_{-\tau}^{\tau} \varphi_{n_k}(t)\, dt \right| \leqq \left| \int_{|x| \leqq b} dF_{n_k}(x) \right| + \frac{1}{b\tau} \left| \int_{|x| > b} dF_{n_k}(x) \right| \leqq a_k + \frac{1}{b\tau}$$
$$\leqq a_k + \frac{\varepsilon}{4} < a + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = a + \frac{\varepsilon}{2}. \tag{6.6.10}$$

Das ist ein Widerspruch zur Ungleichung (6.6.7). Die Funktion $F(x)$ ist also eine Verteilungsfunktion. Aus dem zuerst bewiesenen Grenzwertsatz folgt, daß $\varphi(t)$ ihre charakteristische Funktion ist.

Nicht nur die Teilfolge $\{F_{n_k}(x)\}$, sondern auch die Folge $\{F_n(x)\}$ der Verteilungsfunktionen konvergiert gegen dieselbe Grenzverteilungsfunktion $F(x)$. Wäre das nicht der Fall, so könnte man eine andere Teilfolge $\{F_{n_j}(x)\}$ finden, die gegen eine Grenzfunktion $\tilde{F}(x)$ konvergiert, die nicht identisch gleich $F(x)$ ist. Aus dem Gedankengang unseres Beweises folgt aber, daß $\tilde{F}(x)$ eine Verteilungsfunktion ist, während aus dem Satz 6.6.1 a und aus der Voraussetzung über die Konvergenz der charakteristischen Funktionen folgt, daß $\tilde{F}(x)$ dieselbe charakteristische Funktion wie die Verteilungsfunktion $F(x)$ hat; also besteht nach Satz 4.5.1 die Identität $F(x) \equiv \tilde{F}(x)$. Damit enthält jede Teilfolge von $\{F_n(x)\}$ eine Teilfolge, die gegen dieselbe Verteilungsfunktion $F(x)$ konvergiert; die Folge $\{F_n(x)\}$ konvergiert also gegen $F(x)$.

Aus den Sätzen 6.6.1 a und 6.6.1 b erhalten wir den

Satz 6.6.1 (Lévy-Cramér). *Es sei $\{X_n\}$ $(n = 1, 2, 3, \ldots)$ eine Folge von Zufallsvariablen, und $F_n(x)$ bzw. $\varphi_n(t)$ seien die Verteilungsfunktion bzw. die charakteristische Funktion der Zufallsvariablen X_n. Die Folge $\{F_n(x)\}$ strebt genau dann für $n \to \infty$ gegen die Verteilungsfunktion $F(x)$, wenn in einem gewissen Werteintervall $|t| \leqq \tau$ die Folge $\{\varphi_n(t)\}$ gleichmäßig gegen eine gewisse Funktion $\varphi(t)$ konvergiert. Die Grenzfunktion $\varphi(t)$ ist dann die charakteristische Funktion der Grenzverteilungsfunktion $F(x)$, und die Konvergenz $\varphi_n(t) \to \varphi(t)$ ist gleichmäßig in jedem endlichen Intervall der t-Achse.*

Wir weisen darauf hin, daß der Satz 6.6.1 gültig bleibt, wenn wir voraussetzen, daß die Grenzfunktion $\varphi(t)$ nur an der Stelle $t = 0$ stetig ist (vgl. Cramér [2]). Des weiteren weisen wir darauf hin, daß wir im allgemeinen im Satz 6.6.1 die

Stetigkeit an jeder Stelle t im Intervall $(-\infty, \infty)$ durch die Stetigkeit in einem gewissen, den Punkt $t = 0$ einschließenden Intervall der t-Achse ersetzen können. Sind jedoch sämtliche Variablen X_n wenigstens nach unten oder nach oben gleichmäßig beschränkt, so genügt es für die Konvergenz der Folge $\{F_n(x)\}$ gegen die Verteilungsfunktion $F(x)$, daß in einem gewissen Intervall $|t| < \tau$ die Folge $\{\varphi_n(t)\}$ gegen die an der Stelle $t = 0$ stetige Funktion $\varphi(t)$ konvergiert. Der Beweis hierfür stammt von ZYGMUND [2].

6.7. Der Satz von Moivre-Laplace

A. Zum Beweis des Satzes von MOIVRE-LAPLACE benutzen wir den Satz 6.6.1b.

Mit $\{X_n\}$ bezeichnen wir eine Folge von binomial verteilten Zufallsvariablen. Für jedes n kann die Zufallsvariable X_n die Werte $r = 0, 1, 2, \ldots, n$ annehmen, und ihre Wahrscheinlichkeitsfunktion wird durch

$$P(X_n = r) = \binom{n}{r} p^r q^{n-r}, \text{ wobei } 0 < p < 1 \text{ und } q = 1 - p, \qquad (6.7.1)$$

gegeben. Bekanntlich ist nach den Formeln (5.2.4)

$$E(X_n) = np, \qquad D^2(X_n) = npq.$$

Wir betrachten die Folge $\{Y_n\}$ der normierten Zufallsvariablen

$$Y_n = \frac{X_n - np}{\sqrt{npq}} \qquad (6.7.2)$$

und beweisen einen Grenzwertsatz, der nach MOIVRE und LAPLACE benannt ist.

Satz 6.7.1. *Es sei $\{F_n(y)\}$ die Folge der Verteilungsfunktionen der durch die Formel* (6.7.2) *definierten Zufallsvariablen Y_n. Hierbei sind die in* (6.7.2) *vorkommenden X_n nach der Formel* (6.7.1) *binomial verteilt. Wenn $0 < p < 1$ ist, gilt für jeden Wert von y die Beziehung*

$$\lim_{n\to\infty} F_n(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{y^2}{2}}\, dy. \qquad (6.7.3)$$

Beweis. Die charakteristische Funktion $\varphi_x(t)$ der Zufallsvariablen X_n hat wegen (5.2.3) die Gestalt

$$\varphi_x(t) = (q + pe^{it})^n. \qquad (6.7.4)$$

Nach (4.2.17) wird dann die charakteristische Funktion $\varphi_y(t)$ der Zufallsvariablen Y_n durch

$$\begin{aligned}\varphi_y(t) &= \exp\left(-\frac{npit}{\sqrt{npq}}\right)\left[q + p\exp\left(\frac{it}{\sqrt{npq}}\right)\right]^n \\ &= \left[q\exp\left(-\frac{pit}{\sqrt{npq}}\right) + p\exp\left(\frac{qit}{\sqrt{npq}}\right)\right]^n \qquad (6.7.5)\end{aligned}$$

gegeben.

Wir entwickeln e^{iz} nach der Taylorschen Formel und benutzen die Restformel von PEANO:

$$e^{iz} = \sum_{j=0}^{k} \frac{(iz)^j}{j!} + o(z^k).$$

Daraus erhalten wir

$$p\exp\left(\frac{qit}{\sqrt{npq}}\right) = p + it\sqrt{\frac{pq}{n}} - \frac{qt^2}{2n} + o\left(\frac{t^2}{n}\right),$$

$$q\exp\left(-\frac{pit}{\sqrt{npq}}\right) = q - it\sqrt{\frac{pq}{n}} - \frac{pt^2}{2n} + o\left(\frac{t^2}{n}\right),$$

wobei für jedes t die Beziehung

$$\lim_{n\to\infty} n\,o\left(\frac{t^2}{n}\right) = 0 \qquad (6.7.6)$$

besteht. Wenn wir diese Werte in die Formel (6.7.5) einsetzen und die Gleichung $p + q = 1$ berücksichtigen, erhalten wir

$$\varphi_y(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n.$$

Daraus folgt

$$\log\varphi_y(t) = n\log\left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right] = n\log(1 + z).$$

Da $|z| < 1$ für jedes feste t und genügend große n ist, können wir schreiben:

$$\log\varphi_y(t) = -\frac{t^2}{2} + n\,o\left(\frac{t^2}{n}\right).$$

Wenn wir (6.7.6) berücksichtigen, erhalten wir

$$\lim_{n\to\infty} \log \varphi_y(t) = -\frac{t^2}{2}, \qquad \text{also} \qquad \lim_{n\to\infty} \varphi_y(t) = e^{-\frac{t^2}{2}}.$$

Damit haben wir gezeigt, daß die Folge der charakteristischen Funktionen $\varphi_y(t)$ der normierten Zufallsvariablen Y_n, die durch die Formel (6.7.2) definiert sind, für $n \to \infty$ gegen die charakteristische Funktion (siehe 5.7) einer normalen Zufallsvariablen konvergiert, deren Verteilungsfunktion durch die rechte Seite von (6.7.3) gegeben wird. Aus dem Satz 6.6.1 b erhalten wir sofort die Formel (6.7.3).

Da die Verteilungsfunktion einer normalen Zufallsvariablen keine Unstetigkeitsstellen hat, ist die in der Formel (6.7.3) betrachtete Konvergenz in jedem Punkt y vorhanden. Der Satz von MOIVRE-LAPLACE ist damit bewiesen.

Es seien y_1 und y_2 zwei beliebige Punkte $(y_1 < y_2)$. Aus (6.7.3) folgt

$$\lim_{n\to\infty} P(y_1 < Y_n < y_2) = \lim_{n\to\infty} [F_n(y_2) - F_n(y_1)] = \frac{1}{\sqrt{2\pi}} \int_{y_1}^{y_2} e^{-\frac{y^2}{2}} dy. \quad (6.7.7)$$

Den Satz von MOIVRE-LAPLACE wollen wir noch anders schreiben. Aus (6.7.2) erhalten wir nämlich

$$P(y_1 < Y_n < y_2) = P\left(y_1 < \frac{X_n - np}{\sqrt{npq}} < y_2\right)$$
$$= P\left(y_1\sqrt{npq} + np < X_n < y_2\sqrt{npq} + np\right),$$

also

$$\lim_{n\to\infty} P\left(y_1\sqrt{npq} + np < X_n < y_2\sqrt{npq} + np\right) = \frac{1}{\sqrt{2\pi}} \int_{y_1}^{y_2} e^{-\frac{y^2}{2}} dy.$$

Wir führen die Bezeichnungen

$$x_1 = y_1\sqrt{npq} + np, \qquad x_2 = y_2\sqrt{npq} + np \quad (6.7.8)$$

ein und schreiben die Formel (6.7.7) in der Gestalt

$$P(x_1 < X_n < x_2) \approx \frac{1}{\sqrt{2\pi}} \int_{y_1}^{y_2} e^{-\frac{y^2}{2}} dy,$$

wobei y_1 und y_2 aus den Formeln (6.7.8) bestimmt werden.

Wir sagen, die Zufallsvariable X_n sei *asymptotisch normal* $N(np; \sqrt{npq})$ *verteilt.*

Ersetzen wir y_1 bzw. y_2 durch

$$y_1 + \frac{1}{2\sqrt{npq}} \quad \text{bzw.} \quad y_2 - \frac{1}{2\sqrt{npq}},$$

so erhalten wir eine etwas bessere Näherung.

Beispiel 6.7.1. Wir werfen 100mal ein Geldstück. Dem Erscheinen des Adlers schreiben wir die Zahl Eins, dem Erscheinen des Kopfes die Zahl Null zu. Die Wahrscheinlichkeit jedes Ereignisses beträgt $p = q = 0{,}5$. Wie groß ist die Wahrscheinlichkeit dafür, daß der Adler mehr als 50mal und weniger als 60mal erscheint?

Die Zufallsvariable X_n kann hier die Werte von 0 bis 100 annehmen. Es ist

$$E(X_n) = 50, \quad D^2(X_n) = 25,$$

$$P(50 < X_n < 60) = P\left(\frac{50-50}{5} < \frac{X_n - 50}{5} < \frac{60-50}{5}\right)$$

$$= P\left(0 < \frac{X_n - 50}{5} < 2\right) \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-0,1}^{1,9} e^{-\frac{x^2}{2}}\, dx.$$

Den Tafeln für die Normalverteilung entnehmen wir den Wert dieses Integrals; er beträgt 0,4315.

B. Aus dem Grenzwertsatz von Moivre-Laplace für eine Folge von Zufallsvariablen mit Binomialverteilung erhalten wir einen ähnlichen Grenzwertsatz für die Folge der Zufallsvariablen

$$U_n = \frac{X_n}{n},$$

wobei X_n eine durch die Formel (6.7.1) gegebene Binomialverteilung hat.

Mit Z_n sei die Zufallsvariable bezeichnet, die wir durch Normierung der Zufallsvariablen U_n erhalten. Wenn wir betrachten, daß $E(U_n) = p$ und $D^2(U_n) = \frac{pq}{n}$ ist, erhalten wird die Beziehung

$$Z_n = \frac{\frac{X_n}{n} - p}{\sqrt{\frac{pq}{n}}} = \frac{X_n - np}{\sqrt{npq}} = Y_n,$$

wobei Y_n die durch (6.7.2) definierte Zufallsvariable ist.

Da die Folge der Verteilungsfunktionen $F_n(y)$ der Zufallsvariablen Y_n die Beziehung (6.7.3) erfüllt, gilt für die Folge der Verteilungsfunktionen $F_n(z)$ der

Zufallsvariablen Z_n die Formel

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}}\, dz.$$

Ebenso ergibt sich für beliebige Punkte z_1 und z_2 $(z_1 < z_2)$ die Beziehung

$$\lim_{n\to\infty} P\left(z_1 < \sqrt{n}\, \frac{U_n - p}{\sqrt{pq}} < z_2\right) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}}\, dz. \tag{6.7.9}$$

Wir setzen

$$u_1 = z_1 \sqrt{\frac{pq}{n}} + p, \quad u_2 = z_2 \sqrt{\frac{pq}{n}} + p \tag{6.7.10}$$

und schreiben (6.7.9) in Form der asymptotischen Gleichung

$$P(u_1 < U_n < u_2) \approx \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}}\, dz; \tag{6.7.11}$$

hierbei bestimmen wir die Werte z_1 und z_2 aus der Formel (6.7.10).

Von der Zufallsvariablen U_n, die die Gleichung (6.7.11) erfüllt, sagen wir, daß sie eine *asymptotische Normalverteilung* $N\left(p; \sqrt{\frac{pq}{n}}\right)$ hat.

C. Das in 6.3 bewiesene Bernoullische Gesetz der großen Zahlen erlaubt nur die Feststellung, daß für jedes $\varepsilon > 0$ die Wahrscheinlichkeit der Ungleichung

$$\left| \frac{X_n}{n} - p \right| > \varepsilon$$

für $n \to \infty$ gegen Null strebt. Der hier bewiesene Grenzwertsatz erlaubt uns, näherungsweise (für große n) die Wahrscheinlichkeit dafür zu berechnen, daß der Wert der Zufallsvariablen $\frac{X_n}{n} - p$ im Intervall

$$\left(z_1 \sqrt{\frac{p(1-p)}{n}}, \quad z_2 \sqrt{\frac{p(1-p)}{n}}\right)$$

liegt; z_1 und z_2 sind hier beliebige Zahlen $(z_1 < z_2)$.

Beispiel 6.7.2. In einem Kasten befinde sich eine Anzahl von Karteikarten der Arbeiter eines gewissen Industriezweiges. Unter diesen Werktätigen befinden sich 20% junge und

80% ältere Arbeiter. Wir greifen aufs Geratewohl eine Karteikarte heraus und notieren das auf der Karte angegebene Alter. Vor der Wahl der nächsten Karte legen wir die schon gewählte Karte in den Kasten zurück. Dann beträgt die Wahrscheinlichkeit, die Karte eines Jungarbeiters herauszugreifen, stets 0,2. Wir untersuchen auf diese Weise n Karten. Wie groß muß nun die Zahl n sein, damit mit einer Wahrscheinlichkeit von 0,95 die Häufigkeit einer Karte, die einem Jungarbeiter entspricht, zwischen 0,18 und 0,22 liegt?

Die Häufigkeit einer Karte eines Jungarbeiters wollen wir mit U_n bezeichnen. Wir erhalten

$$E(U_n) = 0{,}2, \qquad D^2(U_n) = \frac{0{,}16}{n}, \qquad \sqrt{D^2(U_n)} = \frac{0{,}4}{\sqrt{n}}$$

und betrachten die Wahrscheinlichkeit

$$P(0{,}18 < U_n < 0{,}22) = P\left(\frac{-0{,}02}{0{,}4/\sqrt{n}} < \frac{U_n - 0{,}2}{0{,}4/\sqrt{n}} < \frac{0{,}02}{0{,}4/\sqrt{n}}\right)$$

$$= P\left(-0{,}05\sqrt{n} < \frac{U_n - 0{,}2}{0{,}4}\sqrt{n} < 0{,}05\sqrt{n}\right) \approx 0{,}95.$$

Aus (6.7.9) ergibt sich

$$0{,}95 \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-0{,}05\sqrt{n}}^{0{,}05\sqrt{n}} e^{-\frac{z^2}{2}}\,dz.$$

In den Tafeln für die Normalverteilung finden wir

$$0{,}05\sqrt{n} \approx 1{,}96, \quad \text{also} \quad n \approx 1537.$$

6.8. Der Satz von Lindeberg-Lévy

A. Der Satz von MOIVRE-LAPLACE ist, wie wir später sehen werden, ein Spezialfall eines allgemeineren Grenzwertsatzes, nämlich des Lindeberg-Lévyschen Satzes.

Wir betrachten die Folge $\{X_k\}$ $(k = 1, 2, \ldots)$ unabhängiger Zufallsvariabler, die alle dieselben Verteilungen und endliche Momente erster und zweiter Ordnung haben mögen. Wir setzen

$$E(X_k) = m, \qquad D^2(X_k) = \sigma^2 \tag{6.8.1}$$

und betrachten die Zufallsvariable Y_n, die folgendermaßen definiert wird:

$$Y_n = X_1 + X_2 + \cdots + X_n. \tag{6.8.2}$$

Es ist

$$E(Y_n) = nm$$

und, da die Zufallsvariablen X_n unabhängig sind,

$$D^2(Y_n) = n\sigma^2.$$

Die Zufallsvariable

$$Z_n = \frac{Y_n - nm}{\sigma\sqrt{n}} \tag{6.8.3}$$

hat dann den Mittelwert 0 und die Standardabweichung 1.

Nun beweisen wir den folgenden Satz von LINDEBERG-LÉVY (vgl. LINDEBERG [1], LÉVY [1]):

Satz 6.8.1. *Haben die unabhängigen Zufallsvariablen $X_1, X_2, \ldots$ alle dieselbe Verteilung und eine endliche Standardabweichung $\sigma \neq 0$, dann erfüllt die Folge der Verteilungsfunktionen $F_n(z)$ der durch die Formeln* (6.8.3) *und* (6.8.2) *definierten Zufallsvariablen Z_n für jeden Wert z die Beziehung*

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz. \tag{6.8.4}$$

Beweis. Wir schreiben die Gleichung (6.8.3) in der Form

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - m).$$

Alle Zufallsvariablen $X_k - m$ haben dieselbe Verteilung und somit auch dieselbe charakteristische Funktion $\varphi_x(t)$. Die charakteristische Funktion $\varphi_z(t)$ der Zufallsvariablen Z_n hat nach den Formeln (4.2.15) und (4.4.3) die Gestalt

$$\varphi_z(t) = \left[\varphi_x\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n. \tag{6.8.5}$$

Wir haben die Existenz des Moments zweiter Ordnung vorausgesetzt, und es ist

$$E(X_k - m) = 0 \quad \text{und} \quad D^2(X_k - m) = \sigma^2,$$

also können wir die Funktion $\varphi_x(t)$ in der Umgebung des Punktes $t = 0$ folgendermaßen nach der Maclaurinschen Formel entwickeln:

$$\varphi_x(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2). \tag{6.8.6}$$

Wenn wir den Ausdruck (6.8.6) in die Formel (6.8.5) einsetzen, erhalten wir

$$\varphi_z(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n;$$

dabei ist für jedes endliche t

$$\lim_{n\to\infty} n o\left(\frac{t^2}{n}\right) = 0. \tag{6.8.7}$$

Wir setzen

$$u = -\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right),$$

dann ergibt sich

$$\log \varphi_z(t) = n \log(1 + u) = n\left[-\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right] = -\frac{t^2}{n} + no\left(\frac{t^2}{n}\right).$$

Wenn wir die Beziehung (6.8.7) beachten, erhalten wir daraus

$$\lim_{n\to\infty} \log \varphi_z(t) = -\frac{t^2}{2},$$

also

$$\lim_{n\to\infty} \varphi_z(t) = e^{-\frac{t^2}{2}}.$$

Der Ausdruck $e^{-\frac{t^2}{2}}$ ist die charakteristische Funktion einer Zufallsvariablen mit einer Gaußschen Verteilung. Wenn wir noch den Satz 6.6.1b benutzen, folgt, daß die Beziehung (6.8.4) erfüllt ist. Der Satz von LINDEBERG-LÉVY ist damit bewiesen.

B. Es seien z_1 und z_2 zwei beliebige Zahlen $(z_1 < z_2)$. Aus (6.8.4) ergibt sich

$$\lim_{n\to\infty} P(z_1 < Z_n < z_2) = \lim_{n\to\infty} [F_n(z_2) - F_n(z_1)] = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz. \tag{6.8.8}$$

Den Satz 6.8.1 formen wir um. Wir erhalten nämlich aus (6.8.3)

$$P(z_1 < Z_n < z_2) = P\left(z_1 < \frac{Y_n - nm}{\sigma\sqrt{n}} < z_2\right)$$
$$= P\left(z_1 \sigma\sqrt{n} + nm < Y_n < z_2 \sigma\sqrt{n} + nm\right).$$

Aus (6.8.8) folgt dann

$$\lim_{n\to\infty} P\left(z_1 < \frac{Y_n - nm}{\sigma\sqrt{n}} < z_2\right) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz. \tag{6.8.9}$$

Wir führen die Bezeichnungen

$$y_1 = z_1\sigma\sqrt{n} + nm, \quad y_2 = z_2\sigma\sqrt{n} + nm \tag{6.8.10}$$

ein und schreiben (6.8.9) in Gestalt der asymptotischen Gleichung

$$P(y_1 < Y_n < y_2) \approx \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz.$$

Hierbei bestimmen wir z_1 und z_2 aus den Formeln (6.8.10). Die durch die Formel (6.8.2) definierte Zufallsvariable Y_n hat also die asymptotische Normalverteilung $N(nm;\ \sigma\sqrt{n})$.

Wenn eine Summe von Zufallsvariablen asymptotisch normal verteilt ist, sagen wir, daß für sie der zentrale Grenzwertsatz gilt; also gilt für die betrachtete Summe Y_n der zentrale Grenzwertsatz.

Beispiel 6.8.1. Nehmen wir an, die Zufallsvariablen $X_k (k = 1, 2, \ldots)$ seien unabhängig, und jede habe eine Zweipunktverteilung, d. h., für jedes k sei

$$P(X_k = 1) = p, \quad P(X_k = 0) = 1 - p \quad (0 < p < 1).$$

Wir betrachten die Grenzverteilung der Zufallsvariablen

$$Y_n = X_1 + X_2 + \cdots + X_n.$$

Wenn man berücksichtigt, daß $E(X_k) = p$ und $D^2(X_k) = pq$ ist, ergibt sich aus dem Satz 6.8.1, daß die Zufallsvariable Y_n im Limes nach $N(np;\ \sqrt{npq})$ normal verteilt ist.

Da die Zufallsvariable Y_n binomial verteilt ist, ist dieses Beispiel eigentlich ein neuer Beweis für den Satz von MOIVRE-LAPLACE, der also, wie wir sehen, ein Spezialfall des Satzes von LINDEBERG-LÉVY ist.

Beispiel 6.8.2. Die Zufallsvariablen $X_n (n = 1, 2, \ldots)$ seien unabhängig und mögen dieselbe Poissonsche Verteilung haben, die durch die Formel

$$P(X_n = r) = \frac{2^r}{r!} e^{-2} \quad (r = 0, 1, 2, \ldots)$$

gegeben wird. Gesucht ist die Wahrscheinlichkeit dafür, daß der Wert der Summe $Y_{100} = X_1 + X_2 + \cdots + X_{100}$ größer als 190 und kleiner als 210 ist.

Die Zufallsvariable Y_{100} hat näherungsweise die Normalverteilung $N(200;\ 10\sqrt{2})$, da jede der Zufallsvariablen X_n die endliche Standardabweichung $\sigma = \sqrt{2}$ und den Mittelwert $m = 2$ hat. Wir erhalten also

$$P(190 < Y_{100} < 210) = P\left(-0{,}709 < \frac{Y_{100} - 200}{10\sqrt{2}} < 0{,}709\right).$$

Aus den Tafeln für die Normalverteilung finden wir für die gesuchte Wahrscheinlichkeit den Wert 0,52.

C. Aus dem Satz von LINDEBERG-LÉVY ergibt sich der

Satz 6.8.2. *Es seien $X_1, X_2, \ldots$ unabhängige Zufallsvariable mit derselben Verteilung, dem Mittelwert m und der Standardabweichung $\sigma \neq 0$, und es sei die Zufallsvariable U_n durch*

$$U_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

definiert. Weiter sei $F_n(v)$ die Verteilungsfunktion der durch

$$V_n = \frac{U_n - E(U_n)}{\sqrt{D^2(U_n)}}$$

definierten Zufallsvariablen V_n. Die Folge der Verteilungsfunktionen $F_n(v)$ erfüllt dann die Beziehung

$$\lim_{n\to\infty} F_n(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v} e^{-\frac{v^2}{2}} dv. \tag{6.8.11}$$

Beweis. Es ist $E(U_n) = m$ und $D^2(U_n) = \frac{\sigma^2}{n}$. Daraus erhalten wir

$$V_n = \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - m}{\frac{\sigma}{\sqrt{n}}} = \frac{\sum_{k=1}^{n} X_k - nm}{\sigma\sqrt{n}} = Z_n,$$

wobei Z_n die durch (6.8.3) definierten Zufallsvariablen sind. Da die Folge der Verteilungsfunktionen $F_n(z)$ die Beziehung (6.8.4) erfüllt, erfüllt die Folge $\{F_n(v)\}$ die Beziehung (6.8.11).

Es seien jetzt v_1 und v_2 $(v_1 < v_2)$ zwei beliebige Zahlen. Dann besteht die Beziehung

$$\lim_{n\to\infty} P(v_1 < V_n < v_2) = \frac{1}{\sqrt{2\pi}} \int_{v_1}^{v_2} e^{-\frac{v^2}{2}} dv. \tag{6.8.12}$$

Wir setzen

$$u_1 = v_1 \frac{\sigma}{\sqrt{n}} + m, \quad u_2 = v_2 \frac{\sigma}{\sqrt{n}} + m. \tag{6.8.13}$$

Die Formel (6.8.12) schreiben wir in Form der asymptotischen Gleichung

$$P(u_1 < U_n < u_2) \approx \frac{1}{\sqrt{2\pi}} \int_{v_1}^{v_2} e^{-\frac{v^2}{2}} dv,$$

wobei wir die Werte v_1 und v_2 aus den Formeln (6.8.13) bestimmen. Die Veränderlichen U_n haben die asymptotische Normalverteilung $N\left(m; \frac{\sigma}{\sqrt{n}}\right)$; also ist das arithmetische Mittel von n unabhängigen Zufallsvariablen mit derselben, aber beliebigen Verteilung, für die wir nur die Existenz des Moments zweiter Ordnung voraussetzen, asymptotisch normal verteilt.

Beispiel 6.8.3. Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien unabhängig und gleichverteilt; ihre Dichte ist also von der Form

$$f(x) = \begin{cases} 1 & \text{für} \quad x \text{ aus dem Intervall } [0, 1], \\ 0 & \text{für} \quad x < 0 \text{ und } x > 1. \end{cases}$$

Nach den Formeln (5.6.4) und (5.6.5) ist

$$m = \frac{1}{2}, \quad \sigma = \frac{1}{\sqrt{12}}.$$

Wir betrachten die Zufallsvariable

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Wegen Satz 6.8.2 sind die Zufallsvariablen Y_n nach $N\left(\frac{1}{2}; \frac{1}{\sqrt{12\,n}}\right)$ asymptotisch normal verteilt. Wir berechnen für $n = 48$ die Wahrscheinlichkeit dafür, daß der Wert der Zufallsvariablen Y_n kleiner als 0,4 ist.

Es ergibt sich[1])

$$P(Y_n < 0{,}4) = P\left(\frac{Y_n - \frac{1}{2}}{\frac{1}{\sqrt{576}}} < \frac{0{,}4 - \frac{1}{2}}{\frac{1}{\sqrt{576}}}\right) = P\left(\frac{Y_n - \frac{1}{2}}{\frac{1}{24}} < -2{,}4\right)$$
$$\approx \Phi(-2{,}4) \approx 0{,}0082.$$

Obwohl die Zufallsvariablen X_k $(k = 1, 2, \ldots)$ im Intervall $[0, 1]$ gleichverteilt sind, hat jedoch, wie wir sehen, ihr arithmetisches Mittel eine Verteilung, in der Werte, die um mehr als ein Zehntel kleiner als $m = 0{,}5$ sind, äußerst selten vorkommen.

Beispiel 6.8.4. Die Zufallsvariablen X_r $(r = 1, 2, \ldots)$ seien unabhängig und mögen dieselbe Verteilung haben. Die Zufallsvariable X_r soll die Werte $k = 0, 1, \ldots, 9$ mit den Wahrscheinlichkeiten $P(X_r = k) = 0{,}1$ (für jedes r) annehmen können.

Es ist

$$m = E(X_r) = \frac{1}{10} \sum_{k=0}^{9} k = 4{,}5,$$

$$\sigma^2 = D^2(X_r) = \frac{1}{10} \sum_{k=0}^{9} (k - m)^2 = \frac{1}{10} \sum_{k=0}^{9} k^2 - m^2 = 28{,}50 - 20{,}25 = 8{,}25,$$

also $\sigma = 2{,}87$.

Wir betrachten die Zufallsvariable

$$Y_{100} = \frac{X_1 + X_2 + \cdots + X_{100}}{100}$$

[1]) Die Verteilungsfunktion einer nach $N(0; 1)$ normal verteilten Zufallsvariablen wollen wir mit Φ bezeichnen.

Wie groß ist die Wahrscheinlichkeit dafür, daß der Wert der Zufallsvariablen X_{100} größer als 5 ist?

Wenn wir den Satz 6.8.2 benutzen, können wir schließen, daß die Zufallsvariable Y_{100} näherungsweise nach $N\left(4{,}5;\ \frac{2{,}87}{10}\right)$ normal verteilt ist. Wir erhalten also

$$P(Y_{100} > 5) = P\left(\frac{Y_{100} - 4{,}5}{0{,}287} > \frac{5 - 4{,}5}{0{,}287}\right) = P\left(\frac{Y_{100} - 4{,}5}{0{,}287} > 1{,}74\right)$$
$$\approx 1 - \Phi(1{,}74) \approx 0{,}041.$$

Ähnliche Fragen, wie sie in diesem Beispiel besprochen wurden, treffen wir häufig in der Statistik bei der Benutzung von Tafeln zufälliger Zahlen an. Von diesen Tafeln wird noch in Kapitel 14 die Rede sein.

D. Wir zeigen an einem Beispiel, daß das arithmetische Mittel von n Zufallsvariablen mit derselben Verteilung nicht asymptotisch normal verteilt zu sein braucht; das ist dann der Fall, wenn kein endliches Moment zweiter Ordnung existiert.

Beispiel 6.8.5. Die Zufallsvariablen X_k $(k = 1, 2, \ldots)$ seien unabhängig und mögen dieselbe Cauchysche Verteilung haben, die durch die Formel (5.10.3) definiert ist. Da die charakteristische Funktion der Zufallsvariablen X_k für jedes k gleich $\varphi_k(t) = e^{-|t|}$ ist, nimmt die charakteristische Funktion $\varphi(t)$ der Zufallsvariablen

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

die Form

$$\varphi(t) = e^{\frac{-n|t|}{n}} = e^{-|t|}$$

an; für ein beliebiges n hat also die Zufallsvariable Y_n eine Cauchysche Verteilung und ist nicht asymptotisch normal verteilt.

Wir erinnern aber daran, daß eine Zufallsvariable mit einer Cauchyschen Verteilung keine endliche Standardabweichung besitzt (siehe 5.10).

E. Es mögen die Zufallsvariablen X_k $(k = 1, 2, 3, \ldots)$ den Voraussetzungen des Satzes 6.8.1 genügen, und es sei $E(X_k) = 0$. Wir betrachten für jedes n die Teilsummen

$$S_j = \sum_{k=1}^{j} X_k \quad (j = 1, 2, \ldots, n).$$

Erdös und Kac [1], [2] haben die Grenzverteilungen der Folgen

$$\left\{\max_{1\le j\le n} \frac{S_j}{\sqrt{n}}\right\}, \quad \left\{\max_{1\le j\le n} \frac{|S_j|}{\sqrt{n}}\right\}, \quad \left\{\frac{1}{n^2} \sum_{j=1}^{n} S_j^2\right\}, \quad \left\{\frac{1}{n^{3/2}} \sum_{j=1}^{n} |S_j|\right\}$$

gefunden. Ihre Artikel stellen den Anfang einer Reihe von fruchtbaren Arbeiten dar, die die Grenzverteilungen einer weiten Klasse von Funktionalen, die auf

den Vektoren $(S_1, \ldots, S_n)$ definiert sind, betreffen. Dabei sind die Voraussetzungen, die die Zufallsvariablen X_k betreffen, bedeutend allgemeiner als diejenigen, die wir angenommen haben. Hier wollen wir diese Ergebnisse nicht weiter behandeln, sondern nur auf die erwähnten Arbeiten von ERDÖS und KAC sowie auf die Arbeiten von DONSKER [1], PROCHOROW [2], [3], SKOROCHOD [1], SPITZER [1], BAXTER und DONSKER [1], VARADARAJAN [1], LAMPERTI [1], BARTOSZYŃSKI [1], [2] und BILLINGSLEY [1] verweisen.

6.9. Der Satz von Ljapunoff

In 6.8 untersuchten wir die Grenzverteilung einer Summe von unabhängigen Zufallsvariablen mit derselben Verteilung und stellten fest: Haben die Zufallsvariablen ein endliches Moment zweiter Ordnung, dann ist ihre Summe asymptotisch normal verteilt. Aber die Verteilung einer Summe von unabhängigen Zufallsvariablen mit verschiedenen Verteilungen braucht nicht gegen eine Normalverteilung zu konvergieren, sogar dann nicht, wenn alle Zufallsvariablen endliche Standardabweichungen haben.

Wir beweisen jetzt einen Satz von LJAPUNOFF [2], der eine hinreichende Bedingung dafür angibt, daß eine Summe von unabhängigen Zufallsvariablen asymptotisch normal verteilt ist.

A. Wir betrachten eine Folge $\{X_k\}$ $(k = 1, 2, \ldots)$ von unabhängigen Zufallsvariablen, die endliche Momente dritter Ordnung haben.

Satz 6.9.1 (LJAPUNOFF). *Es sei $\{X_k\}$ eine Folge von unabhängigen Zufallsvariablen, deren Momente dritter Ordnung existieren, und m_k, $\sigma_k \neq 0$, a_k und b_k seien der Mittelwert, die Standardabweichung, das zentrale Moment dritter Ordnung bzw. das absolute zentrale Moment dritter Ordnung der Zufallsvariablen X_k. Wir setzen*

$$B_n = \sqrt[3]{\sum_{k=1}^{n} b_k}, \qquad C_n = \sqrt{\sum_{k=1}^{n} \sigma_k^2}.$$

Wenn die Bedingung

$$\lim_{n\to\infty} \frac{B_n}{C_n} = 0 \tag{6.9.1}$$

erfüllt ist, genügt die Folge der Verteilungsfunktionen $F_n(z)$ der Zufallsvariablen

$$Z_n = \frac{\sum_{k=1}^{n} (X_k - m_k)}{C_n} \tag{6.9.2}$$

für jedes z der Beziehung

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}}\, dz. \tag{6.9.3}$$

Beweis. Wir setzen

$$Y_k = \frac{X_k - m_k}{C_n}.$$

Mit $\varphi_{x_k}(t)$ sei die charakteristische Funktion der Zufallsvariablen $X_k - m_k$ bezeichnet. Da $E(X_k - m_k) = 0$ ist und die Momente σ_k^2 und a_k existieren, können wir nach Formel (4.3.2) die Funktion $\varphi_{x_k}(t)$ in die Summe

$$\varphi_{x_k}(t) = 1 - \frac{\sigma_k^2 t^2}{2} + \frac{a_k}{6}(it)^3 + o(a_k t^3)$$

entwickeln. Die charakteristische Funktion $\varphi_{y_k}(t)$ der Zufallsvariablen Y_k ist nach Formel (4.2.15) gleich

$$\varphi_{y_k}(t) = \varphi_{x_k}\left(\frac{t}{C_n}\right) = 1 - \frac{\sigma_k^2 t^2}{2C_n^2} + \frac{a_k}{6C_n^3}(it)^3 + o\left(\frac{a_k t^3}{C_n^3}\right) = 1 + u_k.$$

Für jedes endliche t ist

$$\lim_{n\to\infty}\left[o\left(\frac{a_k t^3}{C_n^3}\right) : \frac{a_k t^3}{C_n^3}\right] = 0. \tag{6.9.4}$$

Da im Sinne der Ljapunoffschen Ungleichung (3.4.2) $\sigma_k \leqq \sqrt[3]{b_k}$ ist, folgt aus der Bedingung (6.9.1), daß für jedes endliche t

$$\lim_{n\to\infty}\left|\frac{-\sigma_k^2 t^2}{2C_n^2}\right| \leqq \lim_{n\to\infty}\frac{\sqrt[3]{b_k^2}}{2C_n^2}t^2 \leqq \lim_{n\to\infty}\frac{1}{2}\frac{B_n^2}{C_n^2}t^2 = 0 \tag{6.9.5}$$

ist. Weiter erhalten wir

$$\lim_{n\to\infty}\left|\frac{a_k}{6C_n^3}(it)^3\right| \leqq \lim_{n\to\infty}\frac{b_k}{6C_n^3}|t|^3 \leqq \lim_{n\to\infty}\frac{B_n^3|t|^3}{6C_n^3} = 0. \tag{6.9.6}$$

Aus den Beziehungen (6.9.5) und (6.9.6) folgt

$$\lim_{n\to\infty} u_k = 0;$$

dabei ist die Konvergenz gleichmäßig in bezug auf k. Für jedes t existiert also eine Zahl $N = N(t)$ derart, daß für $n > N$ und alle $k \leqq n$ die Ungleichung $|u_k| < \frac{1}{2}$ besteht. Wir erhalten also

$$\log \varphi_{y_k}(t) = \log(1 + u_k) = \frac{u_k}{1} - \frac{u_k^2}{2} + \frac{u_k^3}{3} - \cdots$$

$$= u_k - \frac{u_k^2}{2}\left(1 - \frac{2}{3}u_k + \frac{2}{4}u_k^2 - \cdots\right) = u_k - \frac{u_k^2}{2}v_k.$$

Es gilt die Ungleichung

$$|v_k| \leqq 1 + \frac{2}{3}|u_k| + \frac{2}{4}|u_k|^2 + \cdots < 1 + |u_k| + |u_k|^2 + \cdots < 1 + \frac{1}{2} + \frac{1}{4} + \cdots = 2.$$

Man kann also

$$\log \varphi_{y_k}(t) = u_k + \vartheta_k u_k^2 \tag{6.9.7}$$

schreiben, wobei $\vartheta_k = -\frac{v_k}{2}$ und $|\vartheta_k| < 1$ ist.

Wir bezeichnen mit $\varphi_z(t)$ die charakteristische Funktion der Zufallsvariablen Z_n. Auf Grund der Formel (4.4.3) ist $\varphi_z(t) = \prod_{k=1}^{n} \varphi_{y_k}(t)$, also

$$\log \varphi_z(t) = \sum_{k=1}^{n} \log \varphi_{y_k}(t).$$

Aus der Gleichung (6.9.7) erhalten wir

$$\log \varphi_z(t) = \sum_{k=1}^{n} (u_k + \vartheta_k u_k^2). \tag{6.9.8}$$

Weiter gilt

$$\sum_{k=1}^{n} u_k = -\frac{t^2}{2} + \sum_{k=1}^{n} \frac{a_k (it)^3}{6C_n^3} + \sum_{k=1}^{n} o\left(\frac{a_k t^3}{C_n^3}\right). \tag{6.9.9}$$

Für jedes endliche t ist

$$\lim_{n\to\infty} \left|\sum_{k=1}^{n} \frac{a_k (it)^3}{6C_n^3}\right| \leqq \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b_k |t^3|}{6C_n^3} = \lim_{n\to\infty} \frac{B_n^3 |t^3|}{6C_n^3} = 0. \tag{6.9.10}$$

Wenn wir die Formel (6.9.4) beachten, erhalten wir also

$$\lim_{n\to\infty} \sum_{k=1}^{n} o\left(\frac{a_k t^3}{C_n^3}\right) = \lim_{n\to\infty} \sum_{k=1}^{n} \left\{\frac{a_k t^3}{6C_n^3} \cdot \left[o\left(\frac{a_k t^3}{C_n^3}\right) : \frac{a_k t^3}{6C_n^3}\right]\right\} = 0. \tag{6.9.11}$$

Nach Berücksichtigung der Beziehungen (6.9.10) und (6.9.11) folgt aus (6.9.9) für jedes endliche t

$$\lim_{n\to\infty} \sum_{k=1}^{n} u_k = -\frac{t^2}{2}. \tag{6.9.12}$$

Wir wollen jetzt den Grenzwert der Summe

$$\sum_{k=1}^{n} u_k^2 = \sum_{k=1}^{n} \left[\frac{-\sigma_k^2 t^2}{2C_n^2} + \frac{a_k(it)^3}{6C_n^3} + o\left(\frac{a_k t^3}{C_n^3}\right)\right]^2$$

bestimmen. Aus der Ljapunoffschen Ungleichung (3.4.2) unter der Bedingung (6.9.1) ergibt sich

$$\begin{aligned}\lim_{n\to\infty} \sum_{k=1}^{n} \frac{\sigma_k^4 t^4}{4C_n^4} &\leqq \lim_{n\to\infty} \sum_{k=1}^{n} \frac{\sqrt[3]{b_k^4}}{4C_n^4}\, t^4 = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b_k \sqrt[3]{b_k}}{4C_n^3 C_n}\, t^4 \\ &\leqq \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b_k}{4C_n^3}\, t^4 = \lim_{n\to\infty} \frac{B_n^3}{4C_n^3}\, t^4 = 0,\end{aligned} \tag{6.9.13}$$

$$\lim_{n\to\infty} \left|\sum_{k=1}^{n} \left[\frac{a_k(it)^3}{6C_n^3}\right]^2\right| \leqq \lim_{n\to\infty} \sum_{k=1}^{n} \left|\frac{a_k t^3}{6C_n^3}\right|^2 \leqq \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b_k^2 t^6}{36C_n^6} \leqq \lim_{n\to\infty} \frac{B_n^6 t^6}{36C_n^6} = 0. \tag{6.9.14}$$

Also ist bei Berücksichtigung der Formel (6.9.4) erst recht

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left[o\left(\frac{a_k t^3}{6C_n^3}\right)\right]^3 = 0. \tag{6.9.15}$$

Ebenso erhalten wir die Beziehungen

$$\lim_{n\to\infty} \left|\sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \frac{\sigma_k^2 a_j i^3 t^5}{6C_n^5}\right| = 0, \tag{6.9.16}$$

$$\lim_{n\to\infty} \left|\sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \frac{t^2\sigma_k^2}{C_n^2}\, o\left(\frac{a_j t^3}{C_n^3}\right)\right| = 0, \tag{6.9.17}$$

$$\lim_{n\to\infty} \left|\sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \frac{a_k i^3 t^3}{3C_n^3}\, o\left(\frac{a_j t^3}{C_n^3}\right)\right| = 0. \tag{6.9.18}$$

Wenn wir die Formeln (6.9.13) bis (6.9.18) berücksichtigen und beachten, daß $|\vartheta_k| < 1$ für jedes $k \leqq n$ ist, so erhalten wir

$$\lim_{n\to\infty} \sum_{k=1}^{n} \vartheta_k u_k^2 = 0. \tag{6.9.19}$$

Aus (6.9.8) folgt, wenn wir die Formeln (6.9.12) und (6.9.19) berücksichtigen, daß für alle t die Beziehung

$$\lim_{n\to\infty} \log \varphi_z(t) = -\frac{t^2}{2}$$

gilt; also ist

$$\lim_{n\to\infty} \varphi_z(t) = e^{-\frac{t^2}{2}}.$$

Aus der letzten Beziehung ergibt sich auf Grund des Satzes 6.6.1b die Formel (6.9.3). Der Satz von LJAPUNOFF ist damit bewiesen.

Falls die Bedingung (6.9.1) erfüllt ist (dies ist oft bei praktischen Problemen der Fall), ist auf Grund des Ljapunoffschen Satzes eine Summe von unabhängigen Zufallsvariablen auch dann asymptotisch normal verteilt, wenn nicht alle Summanden die gleiche Verteilung haben.

Einen anderen Beweis für den Satz von LJAPUNOFF findet der Leser in einer Arbeit von PARZEN [2].

B. Der Ljapunoffsche Satz gibt aber nur eine hinreichende Bedingung dafür an, daß die Beziehung (6.9.3) gilt. Wir führen jetzt ohne Beweis den Satz von LINDEBERG-FELLER an, der eine notwendige und hinreichende Bedingung angibt. Der Leser findet den Beweis in den Originalarbeiten von LINDEBERG [1] und FELLER [2] oder auch im Buch von GNEDENKO [6] oder LOÈVE [4].

Satz 6.9.2 (LINDEBERG-FELLER). *Es sei* $\{X_k\}$ $(k = 1, 2, \ldots)$ *eine Folge von unabhängigen Zufallsvariablen, und* $G_k(x)$, m_k *bzw.* $\sigma_k \neq 0$ *seien die Verteilungsfunktion, der Mittelwert bzw. die Standardabweichung der Zufallsvariablen* X_k. *Weiter sei*

$$C_n = \sqrt{\sum_{k=1}^{n} \sigma_k^2}.$$

Unter diesen Voraussetzungen gilt die Beziehung $\lim\limits_{n\to\infty} \max\limits_{1\le k\le n} \dfrac{\sigma_k}{C_n} = 0$, *und die Folge der Verteilungsfunktionen* $F_n(z)$ *der normierten Zufallsvariablen* Z_n, *die durch die Formel* (6.9.2) *definiert sind, erfüllt die Beziehung*

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}}\, dz$$

genau dann, wenn für jedes $\varepsilon > 0$ *die Beziehung*

$$\lim_{n\to\infty} \frac{1}{C_n^2} \sum_{k=1}^{n} \int_{|x-m_k|>\varepsilon C_n} (x - m_k)^2\, d\, G_k(x) = 0 \qquad (6.9.20)$$

gilt.

Wenn alle Zufallsvariablen X_k stetig sind und $g_k(x)$ die Dichte der Zufallsvariablen X_k ist, nimmt die Bedingung (6.9.20) die Form

$$\lim_{n\to\infty} \frac{1}{C_n^2} \sum_{k=1}^{n} \int_{|x-m_k|>\varepsilon C_n} (x - m_k)^2 g_k(x)\, dx = 0 \tag{6.9.21}$$

an. Wenn dagegen alle Zufallsvariablen X_k diskret sind, die Sprungstellen x_{kl} und die Sprunghöhen p_{kl} $(l = 1, 2, \ldots)$ besitzen, so nimmt die Bedingung (6.9.20) folgende Form an:

$$\lim_{n\to\infty} \frac{1}{C_n^2} \sum_{k=1}^{n} \sum_{|x_{kl}-m_k|>\varepsilon C_n} (x_{kl} - m_k)^2 p_{kl} = 0. \tag{6.9.22}$$

Aus dem Lindeberg-Fellerschen Satz folgt

Satz 6.9.3. *Es sei* $\{X_k\}$ $(k = 1, 2, \ldots)$ *eine Folge von unabhängigen Zufallsvariablen, die gleichmäßig beschränkt sind, d. h., es gibt eine Zahl* $a > 0$ *derart, daß für jedes* k

$$P(|X_k| \leqq a) = 1 \tag{6.9.23}$$

ist, und es sei $D^2(X_k) \neq 0$ *für jedes* k. *Dann gilt die Beziehung* (6.9.3) *genau dann, wenn*

$$\lim_{n\to\infty} C_n^2 = \infty \tag{6.9.24}$$

ist.

Beweis. Es sei (6.9.24) erfüllt. Aus (6.9.23) folgt, daß die Zufallsvariablen $X_k - m_k$ gleichmäßig beschränkt sind; also kann man für jedes $\varepsilon > 0$ ein N derart finden, daß für $n > N$

$$P(|X_k - m_k| < \varepsilon C_n;\ k = 1, 2, \ldots, n) = 1$$

gilt. Daraus folgt sofort (6.9.20).

Es gelte nun die Beziehung (6.9.3), aber (6.9.24) sei nicht erfüllt. Es existiert also ein $C < \infty$ derart, daß

$$\lim_{n\to\infty} C_n^2 = C^2$$

ist. Hieraus und aus den Beziehungen (6.9.3) und (6.9.2) folgt dann, daß $\sum_{k=1}^{\infty} (X_k - m_k)$ eine Normalverteilung $N(0; C)$ besitzt (vgl. Aufgabe 6.16.43). Wir setzen $U = (X_2 - m_2) + (X_3 - m_3) + \cdots$. Die Zufallsvariablen $X_1 - m_1$ und U sind unabhängig, und ihre Summe ist normal verteilt. Nach einem Satz von Cramér ([1], Satz 19) sind beide Komponenten einer solchen Summe normal verteilt. Da aber wegen (6.9.23) die Zufallsvariable $X_1 - m_1$ nicht normal verteilt sein kann,

kann auch die Beziehung (6.9.3) nicht erfüllt sein. Damit ist der Satz 6.9.3 bewiesen.

Aus diesem Satz folgt insbesondere: Die Zufallsvariablen

$$Y_n = \sum_{k=1}^{n} X_k$$

mögen eine verallgemeinerte Binomialverteilung besitzen, d. h., die Zufallsvariablen X_k seien unabhängig, und jede von ihnen habe eine Null-Eins-Verteilung, deren Wahrscheinlichkeitsfunktion durch

$$P(X_k = 1) = p_k, \qquad P(X_k = 0) = q_k = 1 - p_k \qquad (k = 1, 2, \ldots)$$

gegeben wird.

Aus den Formeln (5.1.8) und (5.1.9) ergibt sich hier

$$m_k = E(X_k) = p_k, \qquad \sigma_k^2 = D^2(X_k) = p_k q_k, \qquad C_n^2 = \sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{n} p_k q_k.$$

Die Divergenz der Reihe

$$\sum_{k=1}^{\infty} p_k q_k$$

ist dann notwendig und hinreichend dafür, daß die Zufallsvariable Y_n asymptotisch normal nach

$$N\left(\sum_{k=1}^{n} p_k; \; \sqrt{\sum_{k=1}^{n} p_k q_k}\right)$$

verteilt ist.

Beispiel 6.9.1. Auf einer Baustelle sind Ziegel aus verschiedenen Ziegeleien gestapelt. Man nimmt auf Grund der beobachteten Qualität der zuletzt gelieferten Ziegel an, daß der Ausschuß, d. h. der Anteil der fehlerhaften Ziegel in einer Lieferung, nicht in allen Lieferungen, die aus verschiedenen Ziegeleien stammen, der gleiche ist. Bei der Produktion der i-ten Ziegelei sei p_i der Anteil der fehlerlosen Ziegel. Die Werte von p_i seien

$$p_1 = 0{,}95, \quad p_2 = 0{,}90, \quad p_3 = 0{,}98, \quad p_4 = 0{,}92, \quad p_5 = 0{,}96.$$

Die Wahrscheinlichkeit, einen fehlerlosen Ziegel aus einer Lieferung, die aus der i-ten Ziegelei stammt, herauszugreifen, beträgt p_i. Es sollen alle aus ein- und derselben Ziegelei stammenden Ziegel auf einem Stapel liegen. Wir greifen nun aufs Geratewohl aus jedem Stapel 20 Ziegel heraus. Da die Stapel sehr groß sind, ändert das Wegnehmen von ein paar Ziegeln kaum die Zusammensetzung eines Stapels. Wir können also annehmen, daß die Wahrscheinlichkeit, einen fehlerlosen Ziegel aus dem i-ten Stapel herauszugreifen, konstant bleibt und p_i beträgt. Bei der Qualitätsuntersuchung aller 100 Ziegel zeigte es sich, daß 11 Ziegel Ausschuß waren. Dieses Ergebnis weckte in uns einigen Zweifel, ob unsere Annahmen hinsichtlich der Werte p_i nicht zu optimistisch waren.

Das mathematische Modell dieses Beispiels sieht folgendermaßen aus:

Wir haben 100 unabhängige Zufallsvariable X_k, und jede kann zwei Werte annehmen, den Wert Eins (wenn ein fehlerloser Ziegel vorkommt) und den Wert Null (wenn ein fehlerhafter Ziegel vorkommt). Diese Zufallsvariablen sind in 5 Gruppen eingeteilt. Zur i-ten Gruppe gehören diejenigen Zufallsvariablen, die mit der Wahrscheinlichkeit p_i den Wert 1 annehmen. Wir bilden die Summe

$$Y_{100} = X_1 + \cdots + X_{20} + X_{21} + \cdots + X_{40} + X_{41} + \cdots + X_{60} + X_{61} + \cdots + X_{80} + X_{81} + \cdots + X_{100}.$$

Das ist eine Zufallsvariable mit einer verallgemeinerten Binomialverteilung. Wir erhalten

$$E(Y_{100}) = 20 \cdot 0{,}95 + 20 \cdot 0{,}90 + 20 \cdot 0{,}98 + 20 \cdot 0{,}92 + 20 \cdot 0{,}96 = 94{,}20,$$

$$\begin{aligned} D^2(Y_{100}) &= 20 \cdot 0{,}05 \cdot 0{,}95 + 20 \cdot 0{,}10 \cdot 0{,}90 + 20 \cdot 0{,}02 \cdot 0{,}98 + 20 \cdot 0{,}08 \cdot 0{,}92 \\ &\quad + 20 \cdot 0{,}04 \cdot 0{,}96 \\ &= 5{,}382, \\ \sigma &= 2{,}32. \end{aligned}$$

Bevor wir den zentralen Grenzwertsatz anwenden, müssen wir unsere Aufmerksamkeit dem vorhin erhaltenen Ergebnis zuwenden. Danach war für die Konvergenz einer Zufallsvariablen mit einer verallgemeinerten Binomialverteilung gegen die Normalverteilung notwendig und hinreichend, daß die Reihe $\sum_k p_k q_k$ divergiert. Wenn diese Reihe konvergiert, strebt $p_k q_k$ für $k \to \infty$ gegen Null, also muß $\min(p_k, 1 - p_k) \to 0$ gelten. Die Folge $\{p_k\}$ muß also entweder eine Teilfolge enthalten, die gegen Null konvergiert, oder eine Teilfolge, die gegen Eins konvergiert. Für unser Beispiel bedeutet das: Die Reihe $\sum_k p_k q_k$ konvergiert dann, wenn sehr häufig (theoretisch unendlich oft) Ziegellieferungen ankommen, die entweder nur aus fehlerlosen Ziegeln oder nur aus Ausschuß bestehen. Wie die vieljährige Praxis der Ziegelproduktion und die von uns erhaltene Anzahl fehlerhafter Ziegel zeigt, tritt hier keiner der Fälle ein, für den die Reihe $\sum_k p_k q_k$ konvergiert. Man kann also den zentralen Grenzwertsatz anwenden.

Auf Grund dieses Satzes hat die Zufallsvariable Y_{100} näherungsweise die Normalverteilung $N(94{,}2; 2{,}32)$. Es ist also

$$P(Y_{100} \leqq 89) = P\left(\frac{Y_{100} - 94{,}2}{2{,}32} \leqq -2{,}25\right) \approx \Phi(-2{,}25).$$

Der Tafel für die Normalverteilung entnehmen wir, daß der Wert $\Phi(-2{,}25)$ sehr klein ist, er beträgt ungefähr 0,01. In derartigen Fällen sind wir geneigt zu schließen, daß die für p_i angenommenen Werte doch zu optimistisch waren.

Mit diesem Beispiel sind wir unseren Darlegungen ein wenig vorausgeeilt. Fragen dieser Art werden im zweiten Teil des Buches großen Platz einnehmen und dort systematisch behandelt werden. Das Beispiel brachten wir, um zu zeigen, daß der zentrale Grenzwertsatz nicht nur ein schönes mathematisches Ergebnis ist, sondern auch zur Lösung vieler praktischer Fragen dient.

C. Aus den Betrachtungen von 6.6 bis 6.9 erkennen wir, was für eine hervorragende Rolle die Normalverteilung in der Wahrscheinlichkeitsrechnung und deren Anwendungen spielt.

Nun wollen wir jedoch einen Satz bringen, der zeigt, daß unter ziemlich allgemeinen Voraussetzungen die Folge der Verteilungsfunktionen einer Summe von unabhängigen Zufallsvariablen gegen eine Verteilungsfunktion streben kann, die von der Normalverteilung verschieden ist. Wir erinnern daran, daß wir in 6.1 Arbeiten erwähnten, in denen die Probleme, die mit den Verteilungen von Summen unabhängiger Zufallsvariabler zusammenhängen, in vollem Umfange gelöst werden.

Wir betrachten eine Folge $\{Y_n\}$ $(n = 1, 2, 3, \ldots)$ von Zufallsvariablen, wobei Y_n für jedes n die Summe von n unabhängigen Zufallsvariablen X_{nk} $(k = 1, 2, \ldots, n)$ ist:

$$Y_n = \sum_{k=1}^{n} X_{nk}. \tag{6.9.25}$$

Diese Summen sind allgemeiner als die bis jetzt in diesem Kapitel behandelten Summen. Es war nämlich

$$X_{nk} = X_k \quad (n = 1, 2, \ldots; \quad k = 1, 2, \ldots, n).$$

Die Zufallsvariablen X_{nk} $(k = 1, 2, \ldots, n)$ sollen für jedes n die gleiche Verteilung[1]) haben, die durch die Wahrscheinlichkeitsfunktion

$$P(X_{nk} = x_l) = p_{nl} \qquad (l = 1, 2, \ldots, r) \tag{6.9.26}$$

gegeben sei, wobei

$$0 \leqq p_{nl} \leqq 1, \qquad \sum_{l=1}^{r} p_{nl} = 1$$

und r $(r \geqq 2)$ irgendeine natürliche Zahl ist.

Satz 6.9.4. *Es sei Y_n durch die Formel* (6.9.25) *definiert, und die Zufallsvariablen X_{nk} $(k = 1, 2, \ldots, n)$ seien unabhängig mit derselben durch die Formel* (6.9.26) *gegebenen Verteilung. Ferner bezeichne $F_n(z)$ die Verteilungsfunktion der durch*

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{D^2(Y_n)}}$$

definierten Zufallsvariablen Z_n. Dann gilt:

1. *Wenn die Beziehung*

$$\lim_{n\to\infty} n[p_{n1}p_{n2} + p_{n2}p_{n3} + \cdots + p_{n,r-1}p_{nr}] = \infty \tag{6.9.27}$$

---

[1]) Den Fall, daß die Zufallsvariablen X_{nk} nicht die gleiche Verteilung für alle k aufweisen, hat KUBIK [1] untersucht (siehe Aufgabe 6.16.17).

erfüllt ist, genügt die Folge $\{F_n(z)\}$ *der Beziehung*

$$\lim_{n\to\infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{z^2}{2}\right) dz.$$

2. *Wenn die (endlichen oder unendlichen) Grenzwerte*

$$\lim_{n\to\infty} p_{nl} \quad \textit{und} \quad \lim_{n\to\infty} n p_{nl} \qquad (l = 1, 2, \ldots, r)$$

existieren und die Beziehung

$$\lim_{n\to\infty} n[p_{n1}p_{n2} + p_{n2}p_{n3} + \cdots + p_{n,r-1}p_{nr}] = \lambda \tag{6.9.28}$$

mit $\lambda > 0$ *erfüllt ist, konvergiert die Folge* $\{F_n(z)\}$ *gegen die Verteilungsfunktion einer Zufallsvariablen, die die Summe von* s $(1 \leqq s \leqq r - 1)$ *unabhängigen Zufallsvariablen ist, die entweder eine Poissonsche Verteilung besitzen oder lineare Funktionen von Zufallsvariablen mit Poissonschen Verteilungen sind.*

Dieser Satz wurde von Fisz [2] bewiesen.

In Satz 5.5.1 traten Summen der Form (6.8.25) auf. Es war nämlich Y_n die Anzahl der Erfolge in n Versuchen eines Bernoullischen Schemas, und die Wahrscheinlichkeit p_n eines Erfolges war eine Funktion von n, die der Beziehung

$$\lim_{n\to\infty} n p_n = \lambda \tag{6.9.29}$$

genügte, wobei $\lambda > 0$ war. Wir können dann schreiben:

$$Y_n = \sum_{k=1}^{n} X_{nk};$$

dabei ist X_{nk} gleich der Anzahl der Erfolge (gleich Eins oder Null) beim k-ten Versuch; die X_{nk} sind also unabhängig und haben die gleiche, durch

$$P(X_{nk} = 1) = p_{n1} = p_n,$$
$$P(X_{nk} = 0) = p_{n2} = 1 - p_n$$

gegebene Verteilung. Hier ist $r = 2$. Aus der Formel (6.9.29) folgen die Beziehungen

$$\lim_{n\to\infty} p_{n1} = 0, \quad \lim_{n\to\infty} p_{n2} = 1,$$
$$\lim_{n\to\infty} n p_{n1} = \lambda, \quad \lim_{n\to\infty} n p_{n2} = \infty,$$
$$\lim_{n\to\infty} n p_{n1} p_{n2} = \lambda,$$

wobei $\lambda > 0$ ist. Damit sind alle Voraussetzungen des zweiten Teils von Satz 6.9.4 erfüllt. Die Folge $\{F_n(z)\}$, wobei $F_n(z)$ die Verteilungsfunktion der Zufallsvariablen

$$Z_n = \frac{Y_n - np_n}{\sqrt{np_n(1-p_n)}}$$

ist, strebt also gegen die Verteilungsfunktion einer Zufallsvariablen mit einer Poissonschen Verteilung (der Parameter ist hier gleich λ). Das ist der globale Poissonsche Satz (vgl. 6.1), während der Satz 5.5.1 den lokalen Poissonschen Satz darstellt.

D. Neben der Feststellung, daß die Folge $\{F_n(z)\}$ der Verteilungsfunktionen der durch (6.9.2) gegebenen Zufallsvariablen Z_n gegen die Verteilungsfunktion $\Phi(z)$ der Normalverteilung konvergiert, ist es wichtig abzuschätzen, wie rasch diese Konvergenz erfolgt. Mit anderen Worten, es geht darum, die Differenz $|F_n(z) - \Phi(z)|$ für alle n und alle z abzuschätzen. Der Leser findet die weit vorangetriebene Lösung dieser Probleme für binomial verteilte Z_n in den Arbeiten von USPENSKI [1] und FELLER [11] und für den allgemeinen Fall in den Arbeiten von CRAMÉR [1], [3], ESSEEN [1] und BERRY [1] (siehe Aufgabe 6.16.19).

6.10. Der Satz von Gnedenko

A. In diesem Kapitel haben wir uns bis jetzt nur mit der Konvergenz von Folgen von Verteilungsfunktionen einer Summe von normierten Zufallsvariablen gegen eine Grenzverteilungsfunktion beschäftigt. Im allgemeinen folgt jedoch aus der Konvergenz der Folge der Verteilungsfunktionen gegen eine normale Verteilungsfunktion nicht, daß ein „lokaler Grenzwertsatz" gilt, d. h., die Folge der Dichten (wenn die behandelten Zufallsvariablen stetig sind) bzw. die Folge der Wahrscheinlichkeitsfunktionen (wenn die Zufallsvariablen diskret sind) braucht nicht gegen die Grenzdichte bzw. die Wahrscheinlichkeitsfunktion zu konvergieren (vgl. Aufgabe 6.16.25 und 6.16.26).

Hier wollen wir uns nicht mit der Frage in ihrem ganzen Umfange beschäftigen, sondern nur für besondere Folgen von unabhängigen diskreten Zufallsvariablen einen lokalen Grenzwertsatz beweisen, und zwar für Folgen von unabhängigen diskreten Zufallsvariablen, die alle dieselbe Verteilung haben und nur Werte der Form $a + hk$ annehmen, wobei $h > 0$ ist und k (nicht notwendig alle) ganzzahlige Werte durchläuft. Eine solche Verteilung nennen wir *gitterförmige Verteilung* und die Zahl h die *Schrittweite* der Verteilung. Die Schrittweite h ist offenbar gemeinsamer Teiler der Differenzen der Wertepaare, die die Zufallsvariable X_i annehmen kann. Wir nennen h die *maximale Schrittweite*, wenn h der größte gemeinsame Teiler der genannten Differenzen ist. Zunächst werden wir diejenigen diskreten Zufallsvariablen betrachten, die nur ganzzahlige Werte annehmen können.

Satz 6.10.1 (Gnedenko [5]). *Die unabhängigen diskreten Variablen* X_i $(i = 1, 2, \ldots)$ *mögen alle dieselbe Verteilung haben und mit positiver Wahrscheinlichkeit ganze Werte annehmen, und es sei* $E(X_i) = m$ *sowie* $D^2(X_i) = \sigma^2 > 0$. *Dann ist die Beziehung*

$$\lim_{n\to\infty}\left[\sigma\sqrt{n}P_n(k) - \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z_{nk}^2}{2}\right)\right] = 0 \tag{6.10.1}$$

mit

$$P_n(k) = P\left(\sum_{i=1}^{n} X_i = k\right), \tag{6.10.2}$$

$$z_{nk} = \frac{k - mn}{\sigma\sqrt{n}} \tag{6.10.3}$$

genau dann gleichmäßig bezüglich k *im Intervall* $-\infty < k < \infty$ *erfüllt, wenn die größte Schrittweite der Verteilung der Variablen* X_i *gleich Eins ist.*

Von einer Zufallsvariablen, die die letzte Bedingung erfüllt, sagen wir, daß sie die Bedingung (w) erfüllt.

Ein Spezialfall ist der lokale Grenzwertsatz von Moivre-Laplace für den Fall, daß die Zufallsvariable X_i die Werte Eins bzw. Null mit den Wahrscheinlichkeiten p bzw. $1 - p$ annimmt.

B. Beim Beweis des zuletzt erwähnten Satzes werden wir das folgende Lemma benutzen.

Lemma. *Wenn eine Zufallsvariable* X *nur ganze Werte annimmt und die Bedingung* (w) *erfüllt, dann erfüllt ihre charakteristische Funktion* $\varphi(t)$ *die Bedingungen*

$$\varphi(2\pi) = 1, \qquad |\varphi(t)| < 1 \quad \text{für} \quad 0 < |t| < 2\pi. \tag{6.10.4}$$

Beweis des Lemmas. Es ist

$$\varphi(t) = \sum_{k=-\infty}^{\infty} p_k e^{itk}$$

mit $p_k = P(X = k)$. Daraus ergibt sich

$$\varphi(2\pi) = \sum_{k=-\infty}^{\infty} p_k e^{2\pi ki} = \sum_{k=-\infty}^{\infty} p_k = 1.$$

Es sei nun $0 < |t_0| < 2\pi$ und $|\varphi(t_0)| = 1$. Dann existiert ein a derart, daß $\varphi(t_0) = e^{ia}$ ist:

$$\int_{-\infty}^{\infty} e^{it_0x} dF(x) = e^{ia}.$$

Wenn wir beide Seiten mit e^{-ia} multiplizieren, erhalten wir

$$\int_{-\infty}^{\infty} e^{i(t_0 x - a)}\, dF(x) = 1,$$

also

$$\int_{\infty}^{\infty} \cos(t_0 x - a)\, dF(x) = \sum_{k=-\infty}^{\infty} p_k \cos(t_0 x_k - a) = 1.$$

Diese Bedingung ist nur dann erfüllt, wenn für alle k die Gleichung $\cos(t_0 x_k - a) = 1$ gilt, also wenn

$$x_k = \frac{a}{t_0} + m_k \frac{2\pi}{t_0}, \quad m_k \text{ ganz},$$

ist. Nach Voraussetzung sind die Zahlen x_k ganz. Die Differenzen der Wertepaare der x_k haben also die Form sh, wobei s eine beliebige ganze Zahl und $h = \frac{2\pi}{|t_0|}$ der größte gemeinsame Teiler der Differenzen der x_k ist. Aus der Beziehung $|t_0| < 2\pi$ folgt $h > 1$ im Widerspruch dazu, daß die Bedingung (w) erfüllt ist. Damit ist das Lemma bewiesen.

C. Wir gehen nun zum Beweis des Satzes 6.10.1 über und zeigen zuerst, daß die Bedingung (w) hinreichend ist.

Mit $\varphi(t)$ bzw. $\varphi_n(t)$ seien die charakteristischen Funktionen der Zufallsvariablen X_i bzw. $Y_n = \sum_{i=1}^{n} X_i$ bezeichnet. Es ist

$$\varphi(t) = \sum_{k=-\infty}^{\infty} p_k e^{itk}, \qquad \varphi_n(t) = \sum_{k=-\infty}^{\infty} P_n(k)\, e^{itk}.$$

Auf Grund der Formel (4.5.7) erhalten wir

$$P_n(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_n(t)\, e^{-itk}\, dt,$$

und aus (6.10.3) folgt $k = \sigma\sqrt{n}\, z_{nk} + mn$. Es ergibt sich also

$$P_n(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_n(t) \exp\left[-it\left(\sigma\sqrt{n}\, z_{nk} + mn\right)\right] dt.$$

Wenn wir $u = \sigma\sqrt{n}t$ setzen, finden wir

$$P_n(k) = \frac{1}{2\sigma\sqrt{n}\pi} \int\limits_{-\pi\sigma\sqrt{n}}^{\pi\sigma\sqrt{n}} \varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right) \exp\left[-iu\left(z_{nk} + \frac{mn}{\sigma\sqrt{n}}\right)\right] du. \tag{6.10.5}$$

In Beispiel 4.5.1 leiteten wir die Formel

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-izu-\frac{u^2}{2}} du \tag{6.10.6}$$

her. Wir werden nun zeigen, daß die Differenz

$$R_n = 2\pi\left[\sigma\sqrt{n}P_n(k) - \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z_{nk}^2}{2}\right)\right]$$

im Intervall $-\infty < k < \infty$ gleichmäßig bezüglich k gegen Null konvergiert. Zu diesem Zweck benutzen wir die Formeln (6.10.5) und (6.10.6) und stellen R_n als Summe von vier Integralen dar:

$$R_n = \int\limits_{-\pi\sigma\sqrt{n}}^{\pi\sigma\sqrt{n}} \varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right) \exp\left[-iu\left(z_{nk} + \frac{mn}{\sigma\sqrt{n}}\right)\right] du$$
$$- \int\limits_{-\infty}^{\infty} \exp\left(-iuz_{nk} - \frac{u^2}{2}\right) du = I_1 + I_2 + I_3 + I_4 \tag{6.10.7}$$

mit

$$I_1 = \int\limits_{|u|<A} \exp(-iuz_{nk})\left[\varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right)\exp\left(-\frac{imnu}{\sigma\sqrt{n}}\right) - \exp\left(-\frac{u^2}{2}\right)\right] du,$$

$$I_2 = \int\limits_{A\leq|u|<\varepsilon\sigma\sqrt{n}} \varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right) \exp\left[-iu\left(z_{nk} + \frac{mn}{\sigma\sqrt{n}}\right)\right] du,$$

$$I_3 = \int\limits_{\varepsilon\sigma\sqrt{n}\leq|u|\leq\pi\sigma\sqrt{n}} \varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right) \exp\left[-iu\left(z_{nk} + \frac{mn}{\sigma\sqrt{n}}\right)\right] du,$$

$$I_4 = -\int\limits_{|u|\geq A} \exp\left(-iuz_{nk} - \frac{u^2}{2}\right) du;$$

dabei sind A und ε positive Konstanten, über deren Wahl wir später sprechen werden. Jetzt wollen wir uns mit der Abschätzung dieser Integrale befassen.

a) Da nach Voraussetzung der Zufallsvariablen X_i eine endliche Dispersion haben, folgt aus dem Satz von LINDEBERG-LÉVY (vgl. 6.8) und aus dem Satz 6.6.1a, daß in einem beliebigen endlichen Intervall von u die Beziehung

$$\varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right)\exp\left(-\frac{imn\,u}{\sigma\sqrt{n}}\right)-\exp\left(-\frac{u^2}{2}\right)\to 0$$

gleichmäßig bezüglich u für $n\to\infty$ erfüllt ist. Für ein beliebiges festes A strebt also I_1 gleichmäßig bezüglich z_{nk} für $n\to\infty$ gegen Null.

b) Aus der Voraussetzung über die Existenz des Moments zweiter Ordnung der Zufallsvariablen X_i folgt die Existenz der zweiten Ableitung ihrer charakteristischen Funktionen. In Übereinstimmung mit der Formel (4.3.2) gilt in einer genügend kleinen Umgebung des Punktes $t=0$ die Beziehung

$$\log\left[e^{-imt}\varphi(t)\right]=\log\left[1-\frac{\sigma^2t^2}{2}+o(t^2)\right].$$

Wenn also die Umgebung des Punktes $t=0$ hinreichend klein ist, erhalten wir

$$\log\left[e^{-imt}\varphi(t)\right]=-\frac{\sigma^2t^2}{2}+o(t^2),$$

$$|e^{-imt}\varphi(t)|\leqq e^{-\frac{\sigma^2t^2}{4}}.$$

Damit ist für ein genügend kleines $\varepsilon>0$ im Bereich $|u|<\varepsilon\sigma\sqrt{n}$

$$\left|e^{-\frac{imnu}{\sigma\sqrt{n}}}\varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right)\right|\leqq e^{-\frac{u^2}{4}}.$$

Daraus erhalten wir sofort eine Abschätzung des Integrals I_2:

$$|I_2|\leqq 2\int\limits_A^{\varepsilon\sigma\sqrt{n}}e^{-\frac{u^2}{4}}\,du<2\int\limits_A^{\infty}e^{-\frac{u^2}{4}}\,du=4\sqrt{\pi}[1-\Phi(A)].$$

Wenn wir also A genügend groß wählen, wird $|I_2|$ kleiner als eine beliebig vorgegebene Zahl.

c) Wir schätzen jetzt das Integral I_3 ab. Aus dem oben bewiesenen Lemma folgt, daß für $0<|u|<2\pi\sigma\sqrt{n}$

$$\left|\varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right)\right|<1$$

gilt. Also kann man zu jedem $\varepsilon > 0$ eine solche Zahl $c > 0$ finden, daß für $\varepsilon\sigma\sqrt{n} \leqq |u| \leqq \pi\sigma\sqrt{n}$

$$\left|\varphi_n\left(\frac{u}{\sigma\sqrt{n}}\right)\right| \leqq e^{-nc}$$

gilt. Daraus erhalten wir

$$|I_3| \leqq e^{-nc} \int\limits_{\varepsilon\sigma\sqrt{n}\leqq|u|\leqq\pi\sigma\sqrt{n}} du < 2\pi\sigma\sqrt{n}e^{-nc}.$$

Also konvergiert das Integral I_3 für $n \to \infty$ gleichmäßig bezüglich z_{nk} gegen Null.

d) Die Abschätzung des Integrals I_4 erhalten wir sofort:

$$|I_4| \leqq 2\int\limits_A^\infty e^{-\frac{u^2}{2}}\,du = 2\sqrt{2\pi}[1 - \Phi(A)].$$

Wenn wir also A hinreichend groß wählen, können wir $|I_4|$ kleiner als jede vorgegebene Zahl machen.

Wie wir sehen, konvergieren die Integrale I_1 und I_3 für $n \to \infty$ gleichmäßig bezüglich z_{nk} gegen Null, und zwar unabhängig davon, wie die Werte A und ε gewählt werden. Die Größen $|I_2|$ und $|I_4|$ können durch die Wahl eines genügend großen A und eines hinreichend kleinen ε beliebig klein gemacht werden. Ihre Abschätzungen hängen dabei nicht von n ab. Daraus folgt, daß auch R_n als Summe dieser Integrale für ein genügend großes n beliebig klein gemacht werden kann. Es wurde also bewiesen, daß die Bedingungen dieses Satzes hinreichend sind.

Wir wollen jetzt voraussetzen, daß die Bedingung (w) nicht erfüllt ist, daß also der gemeinsame Teiler h der Differenzen der Wertepaare, die die Zufallsvariable X_i mit positiver Wahrscheinlichkeit annehmen kann, größer als Eins ist. Dann nehmen aber die Zufallsvariablen X_i nur Werte der Form

$$x_k = a + kh \qquad (k = 0, \pm 1, \pm 2, \ldots)$$

mit positiver Wahrscheinlichkeit an; die Menge der Werte, die von den Zufallsvariablen Y_n mit positiver Wahrscheinlichkeit angenommen werden, besteht also aus Zahlen y_k der Form

$$y_k = na + kh \qquad (k = 0, \pm 1, \pm 2, \ldots). \tag{6.10.8}$$

Wir führen folgende Bezeichnung ein:

$$k_n = na + \left[n\,\frac{m-a}{h}\right]h + 1 \qquad (n = 1, 2, \ldots)$$

($[A]$ bezeichnet die größte ganze Zahl nicht größer als A). Dabei kann ohne Beschränkung der Allgemeinheit a so gewählt werden, daß $m - a \neq 0$ gilt. Da die Zahlen k_n nicht von der Form (6.10.8) sind, bestehen die Gleichungen

$$P_n(k_n) = 0 \qquad (n = 1, 2, \ldots). \tag{6.10.9}$$

Aus den Definitionen der Zahlen z_{nk} und k_n folgen die Ungleichungen

$$|z_{nk_n}| \leqq \frac{|h| + 1}{\sigma\sqrt{n}} \qquad (n = 1, 2, \ldots),$$

woraus sich die Beziehung

$$\lim_{n\to\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_{nk_n}^2}{2}\right) = \frac{1}{\sqrt{2\pi}}$$

ergibt. Aus der letzten Beziehung und der Formel (6.10.9) folgt, daß (6.10.1) nicht gleichmäßig bezüglich k gelten kann. Damit ist die Notwendigkeit der Bedingung (w) bewiesen.

D. Es sollen nun die Zufallsvariablen X_i nur Werte der Form $x_k = a + hk$ annehmen, wobei $h > 0$ der größte gemeinsame Teiler der Differenzen der Wertepaare x_k ist. Die Zufallsvariablen

$$Z_i = \frac{X_i - a}{h}$$

nehmen nur ganze Werte an und erfüllen dann die Bedingung (w). Man kann also auf sie den bewiesenen lokalen Grenzwertsatz anwenden.

Eine weitgehende Verallgemeinerung des Gnedenkoschen Satzes hat RICHTER [1] angegeben, während MEJZLER, PARASIUK und RWATSCHEWA [1] sowie RICHTER [2] den lokalen Grenzwertsatz von GNEDENKO auf mehrdimensionale Zufallsvariable verallgemeinerten.

6.11. Die Gesetze der großen Zahlen von Poisson, Tschebyscheff und Chintschin

A. In 6.3 leiteten wir das Bernoullische Gesetz der großen Zahlen her. Dieses Gesetz der großen Zahlen, historisch gesehen das erste, ist aber nur ein Spezialfall von allgemeineren Sätzen, die wir mit dem gemeinsamen Namen *Gesetze der großen Zahlen* bezeichnen wollen.

Wir betrachten zunächst eine Folge von Zufallsvariablen $\{X_k\}$, von denen wir nur voraussetzen, daß für jedes k die Momente erster und zweiter Ordnung

$$E(X_k) = m_k, \quad E[(X_k - m_k)^2] = \sigma_k^2$$

existieren. Die Zufallsvariablen X_k können abhängig oder unabhängig sein. Wegen der Ungleichung (3.3.1) erhalten wir für jedes k und jedes $\varepsilon > 0$

$$P[|X_k - m_k| \geqq \varepsilon] \leqq \frac{\sigma_k^2}{\varepsilon^2}. \tag{6.11.1}$$

Wenn die Markoffsche Bedingung

$$\lim_{k\to\infty} \sigma_k^2 = 0 \tag{6.11.2}$$

erfüllt ist, erhalten wir aus (6.11.1)

$$\lim_{k\to\infty} P[|X_k - m_k| \geqq \varepsilon] = 0.$$

Das ist der Satz von TSCHEBYSCHEFF.

Satz 6.11.1. *Es sei $\{X_k\}$ eine beliebige Folge von Zufallsvariablen mit endlichen Dispersionen σ_k^2. Wenn die Markoffsche Bedingung* (6.11.2) *erfüllt ist, konvergiert die Folge der Zufallsvariablen $X_k - m_k$ stochastisch gegen Null.*

Wir nehmen jetzt an, daß die in diesem Satz betrachteten Zufallsvariablen X_k paarweise unkorreliert sind, und bilden die Zufallsvariablen

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}. \tag{6.11.3}$$

Dann ist

$$E(Y_n) = \frac{1}{n} \sum_{k=1}^{n} m_k.$$

Da die Zufallsvariablen X_k paarweise unkorreliert sind, ist

$$D^2(Y_n) = \frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2.$$

Wenn also

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \sigma_k^2}{n^2} = 0 \tag{6.11.4}$$

ist, folgt aus Satz 6.11.1

$$\lim_{n\to\infty} P[|Y_n - E(Y_n)| \geqq \varepsilon] = 0.$$

Aus dem Tschebyscheffschen Satz erhielten wir also die Folgerung:

Es sei $\{X_k\}$ $(k = 1, 2, \ldots)$ eine Folge von paarweise unkorrelierten Zufallsvaria-

blen, für die die Größen $E(X_k) = m_k$ und $D^2(X_k) = \sigma_k^2$ existieren. Ist die Bedingung (6.11.4) erfüllt, dann konvergiert die Folge von Zufallsvariablen

$$Y_n - \frac{m_1 + m_2 + \cdots + m_n}{n} \qquad (n = 1, 2, \ldots)$$

stochastisch gegen Null.

B. In 5.3 betrachteten wir das Poissonsche Versuchsschema und die damit zusammenhängende verallgemeinerte Binomialverteilung. Wir wollen für dieses Schema die Summe von n unabhängigen Zufallsvariablen X_k $(k = 1, 2, \ldots, n)$ betrachten, die eine Null-Eins-Verteilung mit $P(X_k = 0) = 1 - p_k$, $P(X_k = 1) = p_k$ haben. Da $D^2(X_k) = p_k(1 - p_k) \leqq \frac{1}{4}$ ist, ist also die Bedingung (6.11.4) erfüllt. Die Folgerung aus dem Tschebyscheffschen Satz nimmt dann die folgende Form an, die wir als *Poissonsches Gesetz der großen Zahlen* bezeichnen können:

Satz 6.11.2. *Wenn die Zufallsvariable Y_n das arithmetische Mittel der Zufallsvariablen X_k eines Poissonschen Versuchsschemas ist,*

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

dann konvergiert die Folge

$$\left\{Y_n - \frac{p_1 + p_2 + \cdots + p_n}{n}\right\}$$

stochastisch gegen Null.

C. Wir betrachten jetzt den Fall, daß die paarweise unkorrelierten Zufallsvariablen X_k $(k = 1, 2, \ldots)$ denselben Mittelwert und dieselbe Standardabweichung haben. Für jedes k ist dann

$$E(X_k) = m, \qquad D^2(X_k) = \sigma^2.$$

Wenn wir die durch die Formel (6.11.3) eingeführte Zufallsvariable Y_n benutzen, erhalten wir

$$E(Y_n) = m, \qquad D^2(Y_n) = \frac{\sigma^2}{n}.$$

Es ist also

$$\lim_{n \to \infty} D^2(Y_n) = 0,$$

da σ^2 nach Voraussetzung endlich ist. Nach der Folgerung aus dem Tschebyscheffschen Satz konvergiert die Folge $\{Y_n - m\}$ stochastisch gegen Null. Wir

haben damit das Tschebyscheffsche Gesetz der großen Zahlen (vgl. TSCHEBYSCHEFF [1]) erhalten, das wir folgendermaßen formulieren können:

Satz 6.11.3. *Ist die Zufallsvariable Y_n das arithmetische Mittel von n paarweise unkorrelierten Zufallsvariablen X_k, die denselben Mittelwert und dieselbe Standardabweichung haben, dann konvergiert die Folge $\{Y_n\}$ stochastisch gegen den gemeinsamen Mittelwert m der Zufallsvariablen X_k.*

Wie der Leser bestätigen kann, ist das Bernoullische Gesetz der großen Zahlen ein Spezialfall des Tschebyscheffschen Gesetzes der großen Zahlen.

D. Bei allen bisher betrachteten Gesetzen der großen Zahlen haben wir vorausgesetzt, daß die Zufallsvariablen, deren stochastische Konvergenz untersucht wurde, eine endliche Dispersion besitzen. Wie CHINTSCHIN aber zeigte [1], ist die Existenz einer endlichen Standardabweichung der Zufallsvariablen X_k, wenn sie alle dieselbe Verteilung haben, nicht notwendig dafür, daß die Folge $\{Y_k\}$ ihrer arithmetischen Mittel stochastisch gegen den gemeinsamen Mittelwert m der Zufallsvariablen X_k konvergiert.

Satz 6.11.4 (CHINTSCHIN). *Es sei $\{X_k\}$ $(k = 1, 2, \ldots)$ eine Folge von unabhängigen Zufallsvariablen mit derselben Verteilung und dem endlichen Mittelwert $E(X_k) = m$. Die Folge der Zufallsvariablen*

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \qquad (n = 1, 2, \ldots)$$

konvergiert dann stochastisch gegen den festen Wert m.

Beweis. Es sei $\varphi(t)$ die charakteristische Funktion der Zufallsvariablen X_k. Wegen der Unabhängigkeit der Zufallsvariablen X_k hat die charakteristische Funktion der Zufallsvariablen Y_n die Gestalt

$$\left[\varphi\left(\frac{t}{n}\right)\right]^n. \tag{6.11.5}$$

Da der endliche Mittelwert m existiert, kann man die Funktion $\varphi(t)$ in der Umgebung des Punktes $t = 0$ nach der Maclaurinschen Formel entwickeln:

$$\varphi(t) = 1 + mit + o(t). \tag{6.11.6}$$

Wir setzen den Ausdruck (6.11.6) in (6.11.5) ein und erhalten

$$\left[\varphi\left(\frac{t}{n}\right)\right]^n = \left[1 + \frac{mit}{n} + o\left(\frac{t}{n}\right)\right]^n.$$

Wenn wir ähnlich wie beim Beweis des Moivre-Laplaceschen Satzes vorgehen, erhalten wir

$$\lim_{n\to\infty}\left[\varphi\left(\frac{t}{n}\right)\right]^n = e^{mit}. \tag{6.11.7}$$

Die rechte Seite dieser Formel ist die charakteristische Funktion einer Zufallsvariablen Y mit einer Einpunktverteilung, für die

$$P(Y = m) = 1$$

ist. Nach dem Satz 6.6.1b konvergiert die Folge der Verteilungsfunktionen $F_n(y)$ der Zufallsvariablen Y_n gegen die Verteilungsfunktion der Zufallsvariablen Y; die Folge $\{Y_n\}$ konvergiert also stochastisch gegen den festen Wert m.

Der Satz 6.11.4 ist das *Chintschinsche Gesetz der großen Zahlen.*

Beispiel 6.11.1. Kehren wir zum Beispiel 6.8.5 zurück. Wir zeigten dort, daß die charakteristische Funktion $\varphi(t)$ der Zufallsvariablen

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

wenn die Zufallsvariablen X_k $(k = 1, 2, \ldots, n)$ unabhängig sind und dieselbe Cauchysche Verteilung haben, von der Form $\varphi(t) = e^{-|t|}$ ist.

Für die Folge $\{Y_n\}$ gilt also nicht das Gesetz der großen Zahlen. Dieses Ergebnis ist nicht überraschend, da die Zufallsvariable X_k keinen Mittelwert besitzt (vgl. Aufgabe 6.16.38).

Die notwendigen und hinreichenden Bedingungen für die Gültigkeit des Gesetzes der großen Zahlen für eine Folge unabhängiger Zufallsvariabler haben Kolmogoroff [1] (vgl. Aufgabe 6.16.39) sowie Feller [8] angegeben.

6.12. Das starke Gesetz der großen Zahlen

A. Die bisher betrachteten Gesetze der großen Zahlen besagten, daß unter gewissen Bedingungen die Folge $\{Z_n\}$ der durch

$$Z_n = \frac{1}{n} \sum_{k=1}^{n} X_k - c_n \quad \text{mit} \quad c_n = \frac{1}{n} \sum_{k=1}^{n} E(X_k) \tag{*}$$

definierten Zufallsvariablen stochastisch gegen Null konvergiert. Dabei wurden die X_k als unabhängig vorausgesetzt. Man kann also zu jedem $\varepsilon > 0$ und $\eta > 0$ ein so großes N finden, daß $P(|Z_n| > \varepsilon) < \eta$ für $n \geqq N$ ist. Daraus folgt aber noch nicht, daß man für beliebiges $\varepsilon > 0$ und $\eta > 0$ ein so großes N wählen kann, daß die Relation

$$P\left(\sup_{n \geqq N} |Z_n| > \varepsilon\right) < \eta \tag{6.12.1}$$

gilt. Diese bedeutet nämlich, daß die Wahrscheinlichkeit der Ungleichung $|Z_n| > \varepsilon$ für wenigstens ein n $(n = N, N + 1, \ldots)$ kleiner als η ist; statt der Wahrscheinlichkeit eines Ereignisses $(|Z_n| > \varepsilon)$ haben wir es hier mit der Wahrscheinlichkeit der Summe von Ereignissen

$$(|Z_N| > \varepsilon) \cup (|Z_{N+1}| > \varepsilon) \cup (|Z_{N+2}| > \varepsilon) \cup \cdots$$

zu tun. Man kann zeigen, daß die Beziehung (6.12.1) mit der Beziehung

$$P\left(\lim_{n\to\infty} Z_n = 0\right) = 1 \tag{6.12.2}$$

äquivalent ist. Die letzte Beziehung besagt, daß die Folge $\{Z_n\}$ *fast überall* gegen Null strebt. Diese Äquivalenz beweisen wir am Ende von 6.12.

Die bis jetzt betrachteten Gesetze der großen Zahlen nennen wir *schwache Gesetze der großen Zahlen.*

Definition 6.12.1. Wir sagen, daß für die Folge von Zufallsvariablen X_k $(k = 1, 2, \ldots)$ das *starke Gesetz der großen Zahlen* gültig ist, wenn für die Folge ihrer arithmetischen Mittel Z_n die Beziehung (6.12.1) für beliebiges $\varepsilon > 0$ und $\eta > 0$ gilt.

B. Bisher ist das Problem der notwendigen und hinreichenden Bedingungen für die Gültigkeit des starken Gesetzes der großen Zahlen für eine Folge von unabhängigen Zufallsvariablen noch nicht befriedigend gelöst worden. Ausführliche Informationen über den gegenwärtigen Stand der Forschung auf diesem Gebiet findet man in der Monografie von Loève [1] und in der Arbeit von Chung [1]. Die am weitesten reichenden Ergebnisse wurden von Prochorow [1], [4] erzielt. Wir geben nachstehend den Satz von Kolmogoroff [2] wieder, der notwendige Bedingungen für die Gültigkeit des starken Gesetzes der großen Zahlen angibt. Der Beweis dieses Satzes beruht auf einer von Kolmogoroff [1] bewiesenen Verallgemeinerung der Tschebyscheffschen Ungleichung, die wir nachstehend in 6.12.C betrachten. In 6.12.D werden wir dann das Lemma von Borel-Cantelli (Borel [1], Cantelli [1]) beweisen. Dieses Lemma werden wir dann beim Beweis des nächsten Satzes von Kolmogoroff [7] verwenden, demzufolge für eine Folge von unabhängigen Zufallsvariablen mit derselben Verteilung die Existenz des Erwartungswertes eine notwendige und hinreichende Bedingung für die Gültigkeit des starken Gesetzes der großen Zahlen ist.

Der Abschnitt 6.12.F ist den Fragen der Existenz notwendiger und hinreichender Bedingungen für die Gültigkeit des starken Gesetzes der großen Zahlen in der Terminologie der Varianzen gewidmet. Die entsprechenden Ergebnisse sind dann in den Sätzen 6.12.4 und 6.12.5 formuliert.

C. Die Kolmogoroffsche Ungleichung. *Es seien $X_1, X_2, \ldots, X_n$ unabhängige Zufallsvariable mit endlichen Dispersionen. Ferner sei*

$$U_i = X_i - E(X_i)$$

und

$$Y_k = \sum_{i=1}^{k} U_i \qquad (k = 1, 2, \ldots, n).$$

Dann besteht für jedes $\varepsilon > 0$ *die Ungleichung*

$$P\left(\max_{1 \leqq k \leqq n} |Y_k| \geqq \varepsilon\right) \leqq \frac{1}{\varepsilon^2} D^2(Y_n). \tag{6.12.3}$$

Beweis. Wir bezeichnen mit A_0 das Ereignis, daß $|Y_k| < \varepsilon$ für $k = 1, \ldots, n$ ist, mit A_1 das Ereignis $(|Y_1| \geqq \varepsilon)$ und mit A_j $(j = 2, \ldots, n)$ das Ereignis, welches darin besteht, daß $|Y_k| < \varepsilon$ für $k = 1, 2, \ldots, j-1$ und $|Y_j| \geqq \varepsilon$ ist. Wie wir bestätigen, ist das Ereignis $\left(\max_{1 \leqq k \leqq n} |Y_k| \geqq \varepsilon\right)$ der Summe $\sum_{j=1}^{n} A_j$ äquivalent, also

$$P\left(\max_{1 \leqq k \leqq n} |Y_k| \geqq \varepsilon\right) = P\left(\sum_{j=1}^{n} A_j\right).$$

Da die Ereignisse A_j einander paarweise ausschließen, ist (6.12.3) der Ungleichung

$$\sum_{j=1}^{n} P(A_j) \leqq \frac{1}{\varepsilon^2} D^2(Y_n) \tag{6.12.4}$$

äquivalent. Mit $F_n(y)$ bezeichnen wir die Verteilungsfunktion der Zufallsvariablen Y_n. Dann gilt für jedes y

$$\begin{aligned} F_n(y) = P(Y_n < y) &= P(Y_n < y, A_0) + \cdots + P(Y_n < y, A_n) \\ &= P(A_0) P(Y_n < y \mid A_0) + \cdots + P(A_n) P(Y_n < y \mid A_n) \\ &= P(A_0) F_n(y \mid A_0) + \cdots + P(A_n) F_n(y \mid A_n), \end{aligned}$$

wo $F_n(y \mid A_j)$ die Verteilungsfunktion der bedingten Verteilung der Zufallsvariablen Y_n unter der Bedingung A_j ist. Da $E(Y_n) = 0$ ist, erhalten wir daraus

$$\begin{aligned} D^2(Y_n) &= \int_{-\infty}^{\infty} y^2 \, dF_n(y) = \sum_{j=0}^{n} P(A_j) \int_{-\infty}^{\infty} y^2 \, dF_n(y \mid A_j) \\ &= \sum_{j=0}^{n} P(A_j) \, E(Y_n^2 \mid A_j) \geqq \sum_{j=1}^{n} P(A_j) \, E(Y_n^2 \mid A_j). \end{aligned} \tag{6.12.5}$$

Wir erhalten weiter

$$\begin{aligned} E(Y_n^2 \mid A_j) &= E\left[\left(Y_j^2 + 2 \sum_{k>j} Y_j U_k + \sum_{k>j} U_k^2 + 2 \sum_{k>h>j} U_k U_h\right) \Big| A_j\right] \\ &\geqq E\left[\left(Y_j^2 + 2 \sum_{k>j} Y_j U_k + 2 \sum_{k>h>j} U_k U_h\right) \Big| A_j\right]. \end{aligned}$$

In das Ereignis A_j gehen nur die ersten j Zufallsvariablen U_i ein. Da die Zufallsvariablen X_i unabhängig sind, sind auch die Zufallsvariablen Y_j und U_k für

$k > j$ und die Zufallsvariablen U_k und U_h für $k > h > j$ unabhängig. Daraus ergibt sich für $j \geqq 1$ und für die betrachteten $k, h > j$

$$E(Y_j U_k \mid A_j) = 0, \quad E(U_k U_h \mid A_j) = 0.$$

Wir erhalten also die Ungleichungen

$$E(Y_n^2 \mid A_j) \geqq E(Y_j^2 \mid A_j) \qquad (j = 1, 2, \ldots, n).$$

Aus den Definitionen des Ereignisses A_j und der letzten Ungleichung folgt

$$E(Y_n^2 \mid A_j) \geqq \varepsilon^2 \qquad (j = 1, 2, \ldots, n).$$

Auf Grund der Formel (6.12.5) ergibt sich daraus

$$D^2(Y_n) \geqq \varepsilon^2 \sum_{j=1}^{n} P(A_j).$$

Wir haben die der Kolmogoroffschen Ungleichung äquivalente Formel (6.12.4) erhalten, womit die Kolmogoroffsche Ungleichung bewiesen ist.

Die Kolmogoroffsche Ungleichung wurde (vgl. Aufgabe 6.16.41) von Hájek und Rényi [1] sowie von Birnbaum und Marshall [1] verallgemeinert.

Satz 6.12.1 (Kolmogoroff). *Es sei $\{X_k\}$ eine Folge von unabhängigen Zufallsvariablen mit den Dispersionen $D^2(X_k)$. Wenn die Ungleichung*

$$\sum_{k=1}^{\infty} \frac{D^2(X_k)}{k^2} < \infty \tag{6.12.6}$$

besteht, dann gilt für die Folge $\{X_k\}$ das starke Gesetz der großen Zahlen. Hierbei können wir

$$c_n = \frac{1}{n} \sum_{k=1}^{n} E(X_k)$$

setzen.

Beweis. Es sei

$$Z_n = \sum_{k=1}^{n} \frac{X_k - E(X_k)}{n}.$$

N und m_0 seien gewisse natürliche Zahlen, die durch die Ungleichung

$$2^{m_0} < N \leqq 2^{m_0+1} \tag{6.12.7}$$

miteinander verknüpft sind. Wir schreiben

$$P_N = P\left(\sup_{n \geqq N} |Z_n| > \varepsilon\right)$$

und zeigen, daß man für beliebiges $\varepsilon > 0$ und $\eta > 0$ ein derartiges N wählen kann, daß $P_N < \eta$ ist.

Es sei

$$Q_m = P\left(\max_{2^m < n \leqq 2^{m+1}} |Z_n| > \varepsilon\right),$$

wobei $m \geqq m_0$ eine natürliche Zahl ist. Es besteht die Ungleichung

$$P_N \leqq \sum_{m=m_0}^{\infty} Q_m. \tag{6.12.8}$$

Wenn wir in der Kolmogoroffschen Ungleichung $D^2(X_k)$ durch $D^2\left(\frac{X_k}{2^{m+1}}\right)$ ersetzen, erhalten wir

$$Q_m \leqq P\left(\max_{1 \leqq n \leqq 2^{m+1}} |Z_n| > \varepsilon\right) \leqq \frac{1}{\varepsilon^2 2^{2m+2}} \sum_{k=1}^{2^{m+1}} D^2(X_k).$$

Aus (6.12.8) ergibt sich also

$$P_N \leqq \sum_{m=m_0}^{\infty} \frac{1}{\varepsilon^2 2^{2m+2}} \sum_{k=1}^{2^{m+1}} D^2(X_k).$$

Wenn wir die Reihenfolge der Summation auf der rechten Seite der letzten Formel umkehren, erhalten wir

$$P_N \leqq \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} D^2(X_k) \sum_{m=m(k)}^{\infty} \frac{1}{2^{2m+2}}, \tag{6.12.9}$$

wobei die Zahlen $m(k)$ die Beziehungen

$$m(k) = \begin{cases} m_0 & (1 \leqq k \leqq 2^{m_0+1}), \\ m_0 + j & (2^{m_0+j} + 1 \leqq k \leqq 2^{m_0+j+1};\ j = 1, 2, \ldots) \end{cases}$$

erfüllen. Es ist

$$\sum_{m=m(k)}^{\infty} \frac{1}{2^{2m+2}} = \frac{1}{2^{2m(k)} \cdot 3} \leqq \frac{4}{3k^2},$$

also

$$P_N \leqq \frac{1}{3\varepsilon^2}\left[\frac{1}{2^{2m_0}} \sum_{k=1}^{m_0} D^2(X_k) + 4 \sum_{k=m_0+1}^{\infty} \frac{D^2(X_k)}{k^2}\right]. \tag{6.12.10}$$

Es sei nun $\eta > 0$ eine beliebige positive Zahl. Aus der Bedingung (6.12.6) und der Formel (6.12.7) folgt, daß für ein genügend großes N und damit auch für ein genügend großes m_0

$$\frac{4}{3\varepsilon^2} \sum_{k=m_0+1}^{\infty} \frac{D^2(X_k)}{k^2} < \frac{\eta}{2} \tag{6.12.11}$$

ist. Außerdem erhalten wir noch

$$\frac{1}{3\varepsilon^2}\frac{1}{2^{2m_0}}\sum_{k=1}^{m_0} D^2(X_k) \leqq \frac{1}{3\varepsilon^2 m_0^2}\sum_{k=1}^{m_0} D^2(X_k).$$

Auf Grund des bekannten Kroneckerschen Lemmas folgt aus (6.12.6) die Beziehung

$$\lim_{m_0\to\infty}\frac{1}{m_0^2}\sum_{k=1}^{m_0} D^2(X_k) = 0.$$

Also ist für hinreichend große N

$$\frac{1}{3\varepsilon^2 2^{2m_0}}\sum_{k=1}^{m_0} D^2(X_k) < \frac{\eta}{2}. \tag{6.12.12}$$

Aus den Formeln (6.12.10) bis (6.12.12) erhalten wir

$$P_N < \eta.$$

Der Satz von Kolmogoroff ist damit bewiesen.

Aus dem Satz 6.12.1 ergibt sich nachstehende Folgerung: *Sind die Dispersionen einer Folge von unabhängigen Zufallsvariablen gleichmäßig beschränkt, dann gilt für diese Folge das starke Gesetz der großen Zahlen.* Ist insbesondere X_k die Anzahl der erfolgreichen Versuche (gleich 0 oder 1) im Versuchsschema von Bernoulli, so gilt für die Folge $\{X_k\}$ das starke Gesetz der großen Zahlen, das in diesem Fall das *Borelsche Gesetz der großen Zahlen* genannt wird. Borel entdeckte diesen Satz im Jahre 1909 (Borel [1]). Desgleichen sieht man leicht, daß für die Anzahl der Erfolge im Versuchsschema von Poisson das starke Gesetz der großen Zahlen ebenfalls erfüllt ist.

Dvoretzky [1] fand (vgl. Aufgabe 6.16.30) eine hinreichende Bedingung dafür, daß in einem Versuchsschema, in dem die Ereignisse in beliebiger Weise voneinander abhängig sind, das starke Gesetz der großen Zahlen gilt.

Hinreichende Bedingungen für die Gültigkeit des starken Gesetzes der großen Zahlen in den Termini der Momente von höherer als zweiter Ordnung wurden von Brunk [1] und Prochorow [1], [4] angegeben. Bobroff [1] fand notwendige (jedoch nicht hinreichende) und hinreichende (jedoch nicht notwendige) Bedingungen, wobei sich diese Bedingungen nur wenig voneinander unterscheiden.

D. Nachstehend wollen wir das Lemma von Borel-Cantelli formulieren und beweisen.

Lemma. *Es sei $\{A_n\}$ $(n = 1, 2, \ldots)$ eine Folge zufälliger Ereignisse, und es möge $P(A_n)$ die Wahrscheinlichkeit des Ereignisses A_n bedeuten, wobei $0 < P(A_n) < 1$ $(n = 1, 2, \ldots)$ ist. Dann gelten die folgenden Beziehungen:*

I. *Ist*

$$\sum_{n=1}^{\infty} P(A_n) < \infty, \tag{6.12.13}$$

so treten mit der Wahrscheinlichkeit 1 *nur endlich viele Ereignisse* A_n *ein.*

II. *Sind die Ereignisse* A_n $(n = 1, 2, \ldots)$ *unabhängig und gilt*

$$\sum_{n=1}^{\infty} P(A_n) = \infty, \tag{6.12.14}$$

so treten mit der Wahrscheinlichkeit 1 *unendlich viele Ereignisse* A_n *ein.*

Beweis. Es sei (6.12.13) erfüllt. Wir bezeichnen mit A das zufällige Ereignis, das den Eintritt unendlich vieler Ereignisse A_n bedeutet. Daraus folgt (vgl. Aufgabe 1.8.7)

$$A = \bigcap_{r=1}^{\infty} \bigcup_{n=r}^{\infty} A_n.$$

Da für jedes natürliche r

$$A \subset \bigcup_{n=r}^{\infty} A_n$$

gilt, ergibt sich hieraus (vgl. Aufgabe 1.8.10)

$$P(A) \leqq P\left(\bigcup_{n=r}^{\infty} A_n\right) \leqq \sum_{n=r}^{\infty} P(A_n).$$

Aus (6.12.13) geht hervor, daß für ein hinreichend großes r die rechte Seite der letzten Ungleichung kleiner ist als eine beliebig vorgegebene Zahl $\varepsilon > 0$, und es ist somit $P(A) = 0$. Damit ist die erste Behauptung des Lemmas bewiesen.

Nun seien die Ereignisse A_n unabhängig, und es sei (6.12.14) erfüllt. Das Ereignis $\bar{A}$ als Gegenstück zu A bestehe darin, daß nur endlich viele Ereignisse A_n eintreten. Es ist somit

$$\bar{A} = \bigcup_{r=1}^{\infty} \bigcap_{n=r}^{\infty} \bar{A}_n.$$

Wegen der Unabhängigkeit der Ereignisse A_n haben wir

$$1 - P(A) = P(\bar{A}) = P\left(\bigcup_{r=1}^{\infty} \bigcap_{n=r}^{\infty} \bar{A}_n\right)$$

$$\leqq \sum_{r=1}^{\infty} P\left(\bigcap_{n=r}^{\infty} \bar{A}_n\right) = \sum_{r=1}^{\infty} \prod_{n=r}^{\infty} \left(1 - P(A_n)\right). \tag{6.12.15}$$

Die Beziehung (6.12.14) zieht nach sich, daß für jedes natürliche r das unendliche Produkt auf der rechten Seite von (6.12.15) Null ist, woraus $P(A) = 1$ folgt. Damit ist auch der zweite Teil des Lemmas bewiesen.

Für eine beliebige Folge $\{A_n\}$ zufälliger Ereignisse, d. h. ohne Voraussetzung der Unabhängigkeit dieser Ereignisse, hat LOÈVE [3] notwendige und hinreichende Bedingungen (vgl. Aufgabe 6.16.34a) dafür gefunden, daß die Wahrscheinlichkeit für das Eintreten unendlich vieler Ereignisse A_n gleich 0 ist. NASH [1] fand notwendige und hinreichende Bedingungen (vgl. Aufgabe 6.16.34b) dafür, daß diese Wahrscheinlichkeit gleich 1 ist.

Definition 6.12.2. Wir sagen, daß die Folgen zufälliger Ereignisse $\{X_k\}$ und $\{X_k^*\}$ $(k = 1, 2, \ldots)$ *äquivalent im Chintschinschen Sinne* sind, wenn

$$\sum_{k=1}^{\infty} P(X_k \neq X_k^*) < \infty$$

erfüllt ist.

Satz 6.12.2. *Es seien $\{X_k\}$ und $\{X_k^*\}$ $(k = 1, 2, \ldots)$ zwei im Chintschinschen Sinne äquivalente Folgen zufälliger Ereignisse. Dann gilt das starke Gesetz der großen Zahlen entweder für beide Folgen oder für keine von ihnen.*

Beweis. Es möge A_k das Ereignis $(X_k \neq X_k^*)$ bedeuten. Aus der Definition 6.12.2 und der ersten Behauptung des Lemmas von BOREL-CANTELLI geht hervor, daß gleichzeitig nur endlich viele Ereignisse A_k auftreten können. Für jede Folge (x_k, x_k^*) von Wertepaaren der Zufallsvariablen (X_k, X_k^*) existiert somit ein k_0 derart (für verschiedene Folgen von Wertepaaren (x_k, x_k^*) kann die Zahl k_0 verschieden sein), daß $x_k = x_k^*$ für $k > k_0$ gilt. Da die Ausdrücke

$$\frac{1}{n}\sum_{k=1}^{k_0} x_k, \quad \frac{1}{n}\sum_{k=1}^{k_0} x_k^*$$

für $n \to \infty$ gegen 0 konvergieren, geht aus der Definition 6.12.1 sofort hervor, daß das starke Gesetz der großen Zahlen für die eine der betrachteten Folgen $\{X_k\}$, $\{X_k^*\}$ dann und nur dann gilt, wenn es auch für die andere Folge erfüllt ist

E. Für Zufallsvariable mit derselben Verteilung hat KOLMOGOROFF [7] den nachstehenden Satz angegeben, den man das *Kolmogoroffsche Gesetz der großen Zahlen* nennt.

Satz 6.12.3. *Es sei $\{Y_i\}$ $(i = 1, 2, \ldots)$ eine Folge unabhängiger Zufallsvariabler mit der gleichen Verteilungsfunktion $F(y)$. Eine notwendige und hinreichende Bedingung dafür, daß für eine gewisse Konstante c*

$$P\left(\lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i - c\right] = 0\right) = 1 \tag{6.12.16}$$

gilt, ist die Existenz des Erwartungswertes $E(Y)$ der Variablen Y mit der Verteilungsfunktion $F(y)$, wobei $c = E(Y)$ ist.

Beweis. Es möge $E(Y)$ und damit auch $E(|Y|)$ existieren. Wir setzen dann $X = Y - E(Y)$, so daß $E(X) = 0$ ist. Dann ist

$$\sum_{i=1}^{\infty} P(|X| > i) = \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} P(k < |X| \leqq k+1) \tag{6.12.17}$$

$$= \sum_{k=1}^{\infty} kP(k < |X| \leqq k+1) \leqq E(|X|) < \infty.$$

Wir betrachten die Folge der Zufallsvariablen X_i^*, die für $i = 1, 2, \ldots$ wie folgt definiert sind:

$$X_i^* = \begin{cases} X_i & \text{für} \quad |X_i| \leqq i, \\ 0 & \text{für} \quad |X_i| > i. \end{cases}$$

Dabei ist $X_i = Y_i - E(Y)$. Aus (6.12.17) geht hervor, daß die Folgen $\{X_i\}$ und $\{X_i^*\}$ äquivalent im Chintschinschen Sinne sind. Um daher zu beweisen, daß für $\{X_i\}$ das starke Gesetz der großen Zahlen gilt, genügt es nach Satz 6.12.2 zu zeigen, daß dieses Gesetz für die Folge $\{X_i^*\}$ erfüllt ist.

Zu diesem Zweck zeigen wir, daß für die Folge $\{X_i^*\}$ die Ungleichung (6.12.6) gilt. In der Tat, wir haben

$$D^2(X_i^*) \leqq E(X_i^{*2}) = \int_{-i}^{i} x^2\, dF(x) \leqq \sum_{k=0}^{i} (k+1)^2 P(k < |X| \leqq k+1),$$

woraus

$$\sum_{i=1}^{\infty} \frac{D^2(X_i^*)}{i^2} \leqq \sum_{i=1}^{\infty} \sum_{k=0}^{i} \frac{(k+1)^2}{i^2} P(k < |X| \leqq k+1)$$

$$= P(0 < |X| \leqq 1) \sum_{i=1}^{\infty} \frac{1}{i^2}$$

$$+ \sum_{k=1}^{\infty} (k+1)^2 P(k < |X| \leqq k+1) \sum_{i=k}^{\infty} \frac{1}{i^2}$$

folgt. Wegen

$$\sum_{i=k}^{\infty} \frac{1}{i^2} < \frac{1}{k^2} + \frac{1}{k} < \frac{2}{k}$$

ist

$$\sum_{i=1}^{\infty} \frac{D^2(X_i^*)}{i^2} < 2P(0 < |X| \leqq 1) + 2\sum_{k=1}^{\infty} \frac{(k+1)^2}{k} P(k < |X| \leqq k+1)$$

$$= 2P(0 < |X| \leqq 1) + 2 \sum_{k=1}^{\infty} kP(k < |X| \leqq k+1)$$

$$+ 4 \sum_{k=1}^{\infty} P(k < |X| \leqq k+1) + 2 \sum_{k=1}^{\infty} \frac{P(k < |X| \leqq k+1)}{k}.$$

Offenbar sind der erste, dritte und vierte Summand auf der rechten Seite der letzten Gleichung endlich. Nach Ungleichung (6.12.17) ist auch der zweite Summand endlich; somit ist die Ungleichung (6.12.6) erfüllt. Damit folgt aus dem Satz 6.12.1, daß für die Folge $\{X_i^*\}$ das starke Gesetz der großen Zahlen gilt, und das ist damit für die Folgen $\{X_i\}$ und $\{Y_i\}$ der Fall.

Wir hatten gezeigt, daß die Existenz des Erwartungswertes $E(Y)$ eine hinreichende Bedingung für die Gültigkeit des starken Gesetzes der großen Zahlen für die Folge $\{Y_i\}$ ist. Es existiert somit eine Folge von Konstanten $\{c_n\}$ derart, daß

$$P\left(\lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i - c_n\right] = 0\right) = 1 \tag{6.12.18}$$

ist. Wir müssen noch zeigen, daß $c_n = E(Y)$ gesetzt werden kann. Aus (6.12.18) geht aber hervor, daß die Folge

$$\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i - c_n\right\}$$

stochastisch gegen 0 konvergiert. Da andererseits die Zufallsvariablen Y_i allen Voraussetzungen des Satzes 6.11.4 genügen, konvergiert auch die Folge

$$\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i - E(Y)\right\}$$

stochastisch gegen 0. Hieraus folgt

$$\lim_{n\to\infty} c_n = E(Y). \tag{6.12.19}$$

Aus den Beziehungen (6.12.18) und (6.12.19) folgt die Beziehung (6.12.16), in der $c = E(Y)$ ist.

Nun nehmen wir an, daß für ein gewisses c die Gleichung (6.12.16) erfüllt ist. Wir setzen $U_i = Y_i - c$. Die Gleichung (6.12.16) nimmt nunmehr die Form

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} U_i = 0\right) = 1$$

an, und wir erhalten

$$P\left(\lim_{i\to\infty}\frac{U_i}{i} = 0\right) = 1.$$

Hieraus folgt wiederum, daß höchstens endlich viele Ereignisse $(|U_i| > i)$ eintreten können. Damit ergibt sich aber aus der zweiten Behauptung des Lemmas von Borel-Cantelli sofort die Ungleichung

$$\sum_{i=1}^{\infty} P(|U_i| > i) < \infty. \tag{6.12.20}$$

Wir bezeichnen mit U eine Variable mit der gleichen Verteilung wie U_i. Unter Beachtung von (6.12.20) erhalten wir

$$E(|U|) \leqq \sum_{k=0}^{\infty} (k+1) P(k < |U| \leqq k+1)$$

$$= \sum_{k=0}^{\infty} k P(k < |U| \leqq k+1) + 1 = \sum_{i=1}^{\infty} P(|U| > i) + 1 < \infty.$$

Es existiert also ein Erwartungswert $E(U)$; damit existiert aber auch der Erwartungswert $E(Y)$. Wir können daher, wie oben gezeigt, $c = E(Y)$ setzen.

Damit ist der Satz bewiesen.

Bei der Durchsicht des Beweises für die Beziehung (6.12.20) stellen wir fest, daß die Voraussetzung von der Gleichheit der Verteilung der Zufallsvariablen U_i unwesentlich ist. Wir haben damit gezeigt: Wenn $\{U_i\}$ $(i = 1, 2, \ldots)$ eine Folge unabhängiger Zufallsvariabler ist, für die das starke Gesetz der großen Zahlen für $c_n = 0$ gilt, ist die Ungleichung (6.12.20) erfüllt.

F. Wir geben nun ohne Beweis den Satz von Prochorow [1] an, der notwendige und hinreichende Bedingungen in der Terminologie der Varianzen für die Gültigkeit des starken Gesetzes der großen Zahlen für eine Folge beschränkter Zufallsvariabler X_i formuliert, sofern diese der Bedingung

$$|X_i| = o\left(\frac{i}{\log\log i}\right) \tag{6.12.21}$$

genügen.

Für $r = 1, 2, \ldots$ setzen wir

$$H_r = \frac{1}{2^{2r}} \sum_{i=2^r+1}^{2^{r+1}} D^2(X_i). \tag{6.12.22}$$

Satz 6.12.4 (Prochorow). *Es sei* $\{X_i\}$ $(i = 1, 2, \ldots)$ *eine Folge unabhängiger Zufallsvariabler, die der Beziehung* (6.12.21) *genügen. Dann genügt die Folge* $\{X_i\}$ *genau dann dem starken Gesetz der großen Zahlen, wenn die Ungleichung*

$$\sum_{r=1}^{\infty} \exp\left(-\frac{\varepsilon}{H_r}\right) < \infty \tag{6.12.23}$$

für beliebiges $\varepsilon > 0$ *gilt, wobei* H_r *durch* (6.12.22) *definiert ist.*

Wie jedoch der nachstehende, von Fisz [6] bewiesene Satz (siehe auch Prochorow [5]) besagt, existieren für eine Folge unabhängiger Zufallsvariabler X_i, die beschränkt sind und einer schwächeren Bedingung als (6.12.21), nämlich der Bedingung

$$|X_i| < i \tag{6.12.24}$$

genügen, keine notwendigen und hinreichenden Bedingungen für die Gültigkeit des starken Gesetzes der großen Zahlen, sofern dieses Gesetz in Varianzen ausgedrückt ist.

Satz 6.12.5. *Für eine Folge* $\{X_i\}$ $(i = 1, 2, \ldots)$ *unabhängiger Variabler, die der Beziehung* (6.12.24) *genügen, existieren keine notwendigen und hinreichenden Bedingungen für die Gültigkeit des in Varianzen ausgedrückten starken Gesetzes der großen Zahlen.*

Dieser Satz läßt sich auch noch wie folgt formulieren: *Ist* $W(\sigma_1^2, \sigma_2^2, \ldots)$ *mit* $\sigma_i^2 = D^2(X_i)$ *irgendeine hinreichende Bedingung für die Gültigkeit des starken Gesetzes der großen Zahlen für die Folge* $\{X_i\}$, *so existiert eine Folge unabhängiger Zufallsvariabler, die der Bedingung* (6.12.24) *genügen, für die das starke Gesetz der großen Zahlen gilt und die dennoch die Bedingung* $W(\sigma_1^2, \sigma_2^2, \ldots)$ *nicht erfüllt.*

Beweis. Zum Beweis des Satzes genügt es, zu zeigen, daß zwei Folgen $\{X_i\}$ und $\{Y_i\}$ unabhängiger Zufallsvariabler existieren, die der Bedingung (6.12.24) und der Gleichung

$$D^2(X_i) = D^2(Y_i) = \sigma_i^2$$

genügen, wobei für die eine Folge das starke Gesetz der großen Zahlen gilt, für die andere hingegen nicht.

Wir definieren die Zufallsvariablen X_i und Y_i wie folgt:

$$P\left(X_i = \frac{i}{\log i}\right) = P\left(X_i = -\frac{i}{\log i}\right) = \frac{\log i}{2i}, \tag{6.12.25}$$

$$P(X_i = 0) = 1 - \frac{\log i}{i},$$

$$P(Y_i = \beta i) = P(Y_i = -\beta i) = \frac{1}{2\beta^2 i \log i}, \tag{6.12.26}$$

$$P(Y_i = 0) = 1 - \frac{1}{\beta^2\, i \log i};$$

hierbei ist $0 < \beta < 1$ und $i = 2, 3, \ldots$ Es ist dann

$$D^2(X_i) = D^2(Y_i) = \frac{i}{\log i}.$$

Offenbar erfüllen die Zufallsvariablen X_i die Beziehung (6.12.21). Wir zeigen, daß sie auch der Bedingung (6.12.23) genügen. In der Tat, es ist

$$\exp\left(-\frac{\varepsilon}{H_r}\right) < \exp\left(-\frac{\varepsilon r}{2}\log 2\right) = \frac{1}{2^{\varepsilon r/2}},$$

und für jedes $\varepsilon > 0$ gilt somit

$$\sum_r \exp\left(-\frac{\varepsilon}{H_r}\right) < \sum_r \frac{1}{2^{\varepsilon r/2}} < \infty .$$

Aus dem Satz 6.12.4 geht hervor, daß für die Folge $\{X_i\}$ das starke Gesetz der großen Zahlen erfüllt ist. Für die Folge $\{Y_i\}$ gilt es hingegen nicht. In der Tat, dieses Gesetz sei für $\{Y_i\}$ bei einer beliebigen Folge $\{c_n\}$ erfüllt. Dann ist es auch für $c_n = 0$ $(n = 1, 2, \ldots)$ erfüllt.[1]) Damit gilt aber auch die Beziehung

$$P\left(\lim_{i\to\infty} \frac{Y_i}{i} = 0\right) = 1. \tag{6.12.27}$$

Aus (6.12.26) folgt jedoch

$$\sum_{i=2}^{\infty} P\left(\frac{|Y_i|}{i} = \beta\right) = \infty ,$$

und nach der zweiten Behauptung des Lemmas von BOREL-CANTELLI treten unendlich viele der Ereignisse $(|Y_i| = \beta i)$ ein, was der Beziehung (6.12.27) widerspricht. Das starke Gesetz der großen Zahlen ist also für die Folge $\{Y_i\}$ nicht erfüllt.

Damit ist der Satz 6.12.5 bewiesen.

Es sei darauf hingewiesen, daß man die Nichtexistenz von notwendigen und hinreichenden Bedingungen für die Gültigkeit des in Varianzen ausgedrückten starken Gesetzes der großen Zahlen für eine Folge von Zufallsvariablen, die keinerlei Beschränkungen von der Art (6.12.24) genügen, auch aus den Beispielen hätte herleiten können, die KOLMOGOROFF [2] angegeben hat. Es läßt sich auch zeigen (FISZ [6]), daß in diesem allgemeinen Fall keine notwendigen und hinreichenden, in Momenten beliebiger endlicher Ordnung ausgedrückte Bedingungen existieren.

G. Wir zeigen jetzt, daß die Beziehung (6.12.1) mit der Beziehung (6.12.2) äquivalent ist, und bezeichnen dazu mit A_N das Ereignis, das darin besteht, daß $\sup_{n\geq N} |Z_n| > \varepsilon$ ist $(\varepsilon > 0)$, und mit A das Produkt der Ereignisse A_N:

$$A = \prod_N A_N . \tag{6.12.28}$$

Für jedes N ist

$$A_{N+1} \subset A_N ,$$

[1]) Zu dieser Folgerung kann man auch über einen Satz gelangen, demzufolge für Zufallsvariable Y_i von der Ordnung $O(i)$ und für eine Folge $\{Y_i\}$, für die das starke Gesetz der großen Zahlen gilt, $c_n = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$ gesetzt werden kann.

also besteht nach Satz 1.3.4 die Beziehung

$$P(A) = \lim_{N\to\infty} P(A_N). \tag{6.12.29}$$

Das zu A_N komplementäre Ereignis $\bar{A}_N$ besteht darin, daß für jedes $n \geqq N$ die Beziehung $|Z_n| \leqq \varepsilon$ besteht. Hier gilt für jedes N

$$A_{N+1} \supset \bar{A}_N,$$

also besteht nach Satz 1.3.5 die Beziehung

$$P(\bar{A}) = P\left(\sum_N \bar{A}_N\right) = \lim_{N\to\infty} P(\bar{A}_N). \tag{6.12.29'}$$

Die Beziehung (6.12.1) sei nicht erfüllt. Dann existieren derartige $\varepsilon > 0$ und $\eta > 0$, daß für jedes N

$$P(A_N) \geqq \eta$$

ist. Daraus und aus (6.12.29) erhalten wir die Ungleichung

$$P(A) \geqq \eta > 0, \tag{6.12.30}$$

aus der wiederum folgt, daß die Beziehung (6.12.2) nicht erfüllt ist. Aus (6.12.2) folgt nämlich: Die Wahrscheinlichkeit, daß für jedes N ein derartiges $n \geqq N$ existiert, für das $|Z_n| > \varepsilon$ ist, ist gleich Null; also wäre im Widerspruch zu (6.12.30) $P(A) = 0$.

Jetzt sei die Beziehung (6.12.2) nicht erfüllt. Es gibt dann derartige $\varepsilon > 0$ und $\eta > 0$, für die die Wahrscheinlichkeit, daß für irgendein N das Ereignis $\bar{A}_N$ eintritt, kleiner als $1 - \eta$ ist, oder

$$P(\bar{A}) < 1 - \eta. \tag{6.12.31}$$

Aus der letzten Ungleichung folgt, daß auch die Beziehung (6.12.1) nicht erfüllt ist, denn (6.12.1) besagt, daß es für beliebige $\varepsilon > 0$ und $\eta > 0$ solche N gibt, für die $P(\bar{A}_N) \geqq 1 - \eta$ ist. Wenn wir (6.12.29') berücksichtigen, würden wir $P(\bar{A}) \geqq 1 - \eta$ erhalten. Dies widerspricht aber der Ungleichung (6.12.31).

Die Äquivalenz der Beziehungen (6.12.1) und (6.12.2) ist damit bewiesen.

H. Zum Abschluß geben wir ein Beispiel für eine Folge von Zufallsvariablen, die stochastisch gegen 0 konvergiert, jedoch nicht fast überall gegen 0 konvergent ist (siehe auch Aufgabe 6.16.38).

Beispiel 6.12.1. Wir betrachten die Folge $\{Z_n\}$ $(n = 1, 2, \ldots)$ unabhängiger Zufallsvariabler, wobei

$$P(Z_n = 1) = \frac{1}{n}, \qquad P(Z_n = 0) = 1 - \frac{1}{n} \tag{6.12.32}$$

gilt.

Die Folge $\{Z_n\}$ konvergiert stochastisch gegen 0, denn aus der Gleichung $P(|Z_n| > \varepsilon) = P(Z_n = 1)$, die für $n = 1, 2, \ldots$ bei beliebigem $0 < \varepsilon < 1$ gilt, erhalten wir für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} P(|Z_n| > \varepsilon) \leqq \lim_{n\to\infty} \frac{1}{n} = 0.$$

Die betrachtete Folge $\{Z_n\}$ erfüllt aber die Beziehung (6.12.2) nicht; denn bezeichnen wir mit A_n das Ereignis $(Z_n = 1)$, so folgt aus (6.12.32)

$$\sum_{n=1}^{\infty} P(A_n) = \infty.$$

Aus der Unabhängigkeit der Ereignisse A_n und aus dem Lemma von BOREL-CANTELLI geht hervor, daß die Wahrscheinlichkeit dafür, daß unendlich viele Ereignisse A_n eintreten, gleich 1 ist. Die Wahrscheinlichkeit für die Existenz einer Teilfolge der Folge $\{Z_n\}$, die nicht gegen 0 konvergiert, ist somit gleich 1. Dies widerspricht aber offenbar der Beziehung (6.12.2).

6.13. Mehrdimensionale Grenzverteilungen

Die Sätze 6.6.1a und 6.6.1b, verallgemeinert auf mehrdimensionale Verteilungen, spielen eine grundlegende Rolle in der Theorie der mehrdimensionalen Grenzverteilungen. Wir werden diese Verallgemeinerungen nicht formulieren, da es genügt, in den erwähnten Sätzen die eindimensionalen charakteristischen Funktionen und Verteilungsfunktionen durch die entsprechenden mehrdimensionalen Funktionen zu ersetzen.

Es sei jetzt $(X_1, \ldots, X_r)$ eine Zufallsvariable mit einer durch die Formel (5.12.1′) definierten Polynomialverteilung. Es ist also

$$P(X_1 = k_1, \ldots, X_r = k_r) = \frac{n!}{k_1! \cdots k_r! (n-K)!} \cdot p_1^{k_1} \cdots p_r^{k_r} q^{n-K}, \tag{6.13.1}$$

wobei

$$K = k_1 + \cdots + k_r \quad \text{und} \quad q = 1 - p_1 - \cdots - p_r$$

ist. In 5.12 zeigten wir, daß die charakteristische Funktion

$$\varphi_x(t_1, \ldots, t_r)$$

der Zufallsvariablen $(X_1, \ldots, X_r)$ die Form

$$\varphi_x(t_1, \ldots, t_r) = \left(\sum_{j=1}^{r} p_j e^{it_j} + q\right)^n \tag{6.13.2}$$

hat, und bewiesen die Gleichungen

$$\begin{aligned} E(X_j) &= np_j, \\ \lambda_{jj} &= D^2(X_j) = np_j(1-p_j) \qquad (j = 1, 2, \ldots, r), \\ \lambda_{jk} &= E[(X_j - np_j)(X_k - np_k)] \\ &= -np_j p_k \qquad (j, k = 1, 2, \ldots, r;\ j \neq k). \end{aligned} \tag{6.13.3}$$

Wir betrachten jetzt die Zufallsvariable $(Z_1, \ldots, Z_r)$; dabei ist

$$Z_j = \frac{X_j - np_j}{\sqrt{np_j(1-p_j)}} \qquad (j = 1, 2, \ldots, r). \tag{6.13.4}$$

Die Momente η_{jk} zweiter Ordnung der Zufallsvariablen Z_j sind durch die Formeln

$$\begin{aligned} \eta_{jj} &= 1 \qquad (j = 1, \ldots, r), \\ \eta_{jk} &= -\sqrt{\frac{p_j p_k}{(1-p_j)(1-p_k)}} \qquad (j, k = 1, \ldots, r;\ j \neq k) \end{aligned} \tag{6.13.4'}$$

gegeben.

Der folgende Satz ist eine Verallgemeinerung des Moivre-Laplaceschen Satzes auf Polynomialverteilungen.

Satz 6.13.1. *Die Zufallsvariable* $(X_1, \ldots, X_r)$ *habe eine durch die Formel* (6.13.1) *gegebene Polynomialverteilung, und* $F_n(z_1, z_2, \ldots, z_r)$ *sei die Verteilungsfunktion der durch die Formel* (6.13.4) *definierten Zufallsvariablen* $(Z_1, \ldots, Z_r)$. *Ist* $0 < q < 1$ *und* $0 < p_j < 1$ $(j = 1, 2, \ldots, r)$, *dann genügen die Glieder der Folge* $\{F_n(z_1, \ldots, z_r)\}$ *der Beziehung*

$$\lim_{n\to\infty} F_n(z_1, \ldots, z_r) = \frac{1}{(2\pi)^{\frac{r}{2}}\sqrt{|N|}} \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_r} \exp\left(-\frac{1}{2|N|}\sum_{j,k=1}^{r} |N_{jk}| z_j z_k\right) dz_1 \cdots dz_r; \tag{6.13.5}$$

dabei ist N die Momentenmatrix, deren Elemente die η_{jk} *sind, und*

$$\begin{aligned} |N| &= \frac{q}{(1-p_1)\cdots(1-p_r)} \neq 0, \\ |N_{jj}| &= \frac{(1-p_j)(q+p_j)}{(1-p_1)\cdots(1-p_r)} \qquad (j = 1, 2, \ldots, r), \\ |N_{jk}| &= \frac{\sqrt{p_j p_k (1-p_j)(1-p_k)}}{(1-p_1)\cdots(1-p_r)} \qquad (j, k = 1, 2, \ldots, r;\ k \neq j). \end{aligned} \tag{6.13.6}$$

Beweis. Wir bezeichnen mit $\varphi_z(t_1, \ldots, t_r)$ die charakteristische Funktion der Zufallsvariablen $(Z_1, \ldots, Z_r)$. Aus den Formeln (6.13.3) und (6.13.4) erhalten wir

$$\varphi_z(t_1, \ldots, t_r) = \exp\left(-\sum_{j=1}^{r} it_j \sqrt{\frac{np_j}{1-p_j}}\right)\left\{\sum_{j=1}^{r} p_j \exp\left[\frac{it_j}{\sqrt{np_j(1-p_j)}}\right] + q\right\}^n,$$

$$\log \varphi_z(t_1, \ldots, t_r) = -i\sum_{j=1}^{r} t_j \sqrt{\frac{np_j}{1-p_j}} + n \log\left\{\sum_{j=1}^{r} p_j \exp\left[\frac{it_j}{\sqrt{np_j(1-p_j)}}\right] + q\right\}$$

$$= -i\sqrt{n}\sum_{j=1}^{r} t_j \sqrt{\frac{p_j}{1-p_j}}$$

$$+ n \log\left[1 + \frac{i}{\sqrt{n}}\sum_{j=1}^{r} t_j \sqrt{\frac{p_j}{1-p_j}} - \frac{1}{2n}\sum_{j=1}^{r} \frac{t_j^2}{1-p_j} + o\left(\frac{1}{n}\right)\right]$$

$$= -i\sqrt{n}\sum_{j=1}^{r} t_j \sqrt{\frac{p_j}{1-p_j}} + n \log(1+s).$$

Dabei haben wir mit s den Ausdruck

$$\frac{i}{\sqrt{n}}\sum_{j=1}^{r} t_j \sqrt{\frac{p_j}{1-p_j}} - \frac{1}{2n}\sum_{j=1}^{r} \frac{t_j^2}{1-p_j} + o\left(\frac{1}{n}\right)$$

bezeichnet. Wenn wir berücksichtigen, daß $|s| < 1$ für beliebige feste $t_1, \ldots, t_r$ und genügend großes n ist, so können wir schreiben:

$$\lim_{n\to\infty}\left[n \log(1+s) - i\sqrt{n}\sum_{j=1}^{r} t_j \sqrt{\frac{p_j}{1-p_j}}\right]$$

$$= -\frac{1}{2}\left[\sum_{j=1}^{r} \frac{t_j^2}{1-p_j} - \left(\sum_{j=1}^{r} t_j \sqrt{\frac{p_j}{1-p_j}}\right)^2\right]$$

$$= -\frac{1}{2}\left[\sum_{j=1}^{r} t_j^2 - \sum_{\substack{j,k=1 \\ j\neq k}}^{r} t_j t_k \sqrt{\frac{p_j p_k}{(1-p_j)(1-p_k)}}\right].$$

Endgültig erhalten wir also

$$\lim_{n\to\infty} \varphi_z(t_1, \ldots, t_r) = \exp\left\{-\frac{1}{2}\left[\sum_{j=1}^{r} t_j^2 - \sum_{\substack{j,k=1 \\ j\neq k}}^{r} t_j t_k \sqrt{\frac{p_j p_k}{(1-p_j)(1-p_k)}}\right]\right\}. \tag{6.13.7}$$

Wir vergleichen (6.13.7) mit (5.11.9) und sehen, daß auf der rechten Seite von (6.13.7) die charakteristische Funktion der r-dimensionalen normalen Zufallsvariablen steht, deren Momentenmatrix durch die Gleichungen (6.13.4′) gegeben ist. Wenn wir endlich die Verallgemeinerung des Satzes 6.6.1b berücksichtigen, von der am Anfang dieses Paragraphen die Rede war, erhalten wir die Formel (6.13.5).[1] Die Formeln (6.13.6) lassen sich leicht ausrechnen.

In 5.12 zeigten wir, daß eine polynomial verteilte Zufallsvariable $(X_1, \ldots, X_r)$, deren Verteilung durch die Formel (6.13.1) gegeben ist, als Summe von n unabhängigen r-dimensionalen Zufallsvariablen mit gleicher Verteilung dargestellt werden kann. Von diesem Standpunkt aus kann man den Satz 6.13.1 als Grenzwertsatz für eine Folge von Verteilungsfunktionen von Summen unabhängiger r-dimensionaler Zufallsvariabler mit gleichen Verteilungen betrachten.

In der Tat läßt sich der Lindeberg-Lévysche Satz auf mehrdimensionale Zufallsvariable verallgemeinern.

Wir wollen diesen Satz hier nur formulieren:

Satz 6.13.2. *Es sei* $\{(Y_{m1}, Y_{m2}, \ldots, Y_{mr})\}$ $(m = 1, 2, 3, \ldots)$ *eine Folge von unabhängigen r-dimensionalen Zufallsvektoren mit gleicher Verteilung und der Momentenmatrix* $M = (\lambda_{jk})$ *mit von Null verschiedener Determinante* $|M|$. *Weiter sei* $F_n(z_1, \ldots, z_r)$ *die Verteilungsfunktion der Zufallsvariablen* $(Z_{n1}, \ldots, Z_{nr})$, *wobei*

$$Z_{nj} = \sum_{m=1}^{n} \frac{Y_{mj} - E(Y_{mj})}{\sqrt{n\,\lambda_{jj}}} \qquad (j = 1, 2, \ldots, r).$$

Die Glieder der Folge $\{F_n(z_1, \ldots, z_r)\}$ *erfüllen dann die Beziehung*

$$\lim_{n\to\infty} F_n(z_1, \ldots, z_r)$$

$$= \frac{1}{(2\pi)^{\frac{r}{2}} \sqrt{|N|}} \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_r} \exp\left(-\frac{1}{2|N|} \sum_{j,k=1}^{r} |N_{jk}| z_j z_k\right) dz_1 \cdots dz_r\,;$$

dabei ist $N = (\eta_{jk})$ *die Momentenmatrix, deren Elemente* η_{jk} *durch*

$$\eta_{jj} = 1 \qquad (j = 1, 2, \ldots, r),$$

$$\eta_{jk} = \frac{\lambda_{jk}}{\sqrt{\lambda_{jj}\lambda_{kk}}} \qquad (j, k = 1, 2, \ldots, r;\ j \neq k)$$

gegeben sind.

---

[1]) Auch den Poissonschen Satz kann man auf die Polynomialverteilung verallgemeinern. Fisz [3] gab die Klasse aller möglichen Grenzverteilungen für die Polynomialverteilung an (vgl. Aufgabe 6.16.45).

Wir überlassen dem Leser die Durchführung dieses Beweises, der übrigens vollständig analog zum Beweis des Satzes 6.3.1 verläuft.

Eine Verallgemeinerung des Satzes von LJAPUNOFF auf mehrdimensionale Verteilungen hat BERNSTEIN [3] angegeben. Den auf mehrdimensionale Verteilungen verallgemeinerten Satz von LINDEBERG-FELLER findet man in einer Arbeit von GNEDENKO [9]. MOURIER [1] hat eine weitreichende Verallgemeinerung des Satzes von LINDEBERG-LÉVY für Zufallsvariable gegeben, die Werte in einem Hilbertraum durchlaufen.

6.14. Grenzwertsätze für rationale Funktionen von Zufallsvariablen

A. Wir bringen noch einige Sätze, die eine große Rolle in den Anwendungen spielen.

Satz 6.14.1. *Es sei $\{X_n\}$ eine beliebige Folge von abhängigen oder unabhängigen Zufallsvariablen, und es konvergiere die entsprechende Folge der Verteilungsfunktionen $F_n(x)$ für $n \to \infty$ gegen die Verteilungsfunktion $F(x)$.*

Weiter sei $\{Y_n\}$ eine beliebige andere Folge von Zufallsvariablen, die stochastisch gegen die Konstante a konvergiert. Dann konvergiert die Folge der Verteilungsfunktionen

α) *der Zufallsvariablen $X_n + Y_n$ gegen die Verteilungsfunktion $F(x - a)$,*

β) *der Zufallsvariablen $X_n - Y_n$ gegen die Verteilungsfunktion $F(x + a)$,*

γ) *der Zufallsvariablen $X_n Y_n$ gegen die Verteilungsfunktion $F\left(\frac{x}{a}\right)$, falls $a > 0$,*

und gegen die Verteilungsfunktion $1 - F\left(\frac{x}{a}\right)$, falls $a < 0$ ist,

δ) *der Zufallsvariablen $\frac{X_n}{Y_n}$ gegen die Verteilungsfunktion $F(ax)$, falls $a > 0$,*

und gegen die Verteilungsfunktion $1 - F(ax)$, falls $a < 0$ ist.

Wir beschränken uns auf den Beweis der Behauptung γ) für $a > 0$, da die Beweise der übrigen Behauptungen analog sind.

Mit $\Phi_n(x)$ bezeichnen wir die Verteilungsfunktion des Produkts $X_n Y_n$. Es sei $\frac{x}{a}$ eine Stetigkeitsstelle der Grenzverteilungsfunktion $F(x)$. Wir haben die Beziehung

$$\Phi_n(x) = P(X_n Y_n < x) \to F\left(\frac{x}{a}\right) \tag{6.14.1}$$

zu beweisen. Mit S bezeichnen wir die Menge aller Punkte in der Ebene, für die die Ungleichung $x_n y_n < x$ erfüllt ist. Man kann S auch als Vereinigung von zwei

punktfremden Mengen S_1 und S_2 darstellen, denen die Punkte angehören, die die Ungleichungen

$$x_n y_n < x, \quad |y_n - a| \leqq \varepsilon \tag{6.14.2}$$

bzw.

$$x_n y_n < x, \quad |y_n - a| > \varepsilon \tag{6.14.3}$$

erfüllen, wobei $0 < \varepsilon < a$ ist. Wir bezeichnen weiter mit A bzw. B diejenigen Mengen in der x_n, y_n-Ebene, die durch die Ungleichungen

$$x_n(a - \varepsilon) < x, \quad |y_n - a| \leqq \varepsilon \tag{6.14.4}$$

bzw.

$$x_n(a + \varepsilon) < x, \quad |y_n - a| \leqq \varepsilon \tag{6.14.5}$$

definiert sind.

In Abb. 6.14.1 (sie stellt nur die Halbebene $y_n > 0$ dar) liegt die Menge S außerhalb der Hyperbel $x_n y_n = x$. Die Menge S_1 wurde schraffiert, die Menge $A - S_1$ mit dem Symbol I und die Menge $S_1 - B$ mit dem Symbol II bezeichnet.

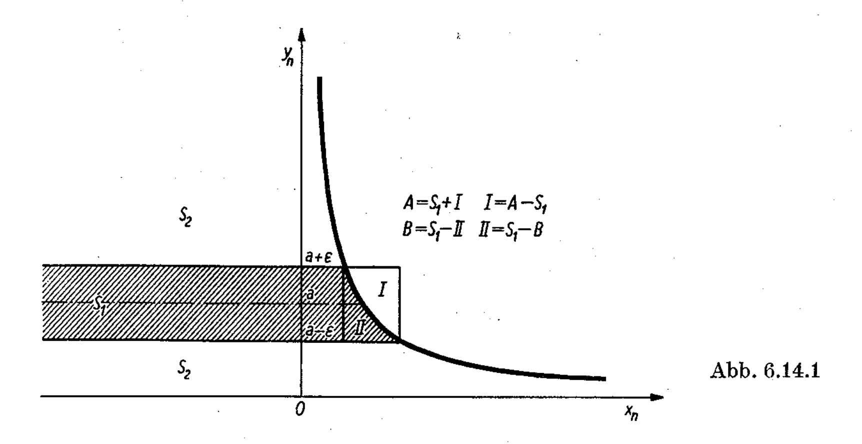


Abb. 6.14.1

Wir bezeichnen mit $P_n(S)$, $P_n(S_1)$, $P_n(S_2)$, $P_n(A)$ und $P_n(B)$ die entsprechenden Wahrscheinlichkeiten dafür, daß die zweidimensionale Zufallsvariable (X_n, Y_n) Werte aus diesen Mengen annimmt.

Es ist

$$\Phi_n(x) = P(X_n Y_n < x) = P_n(S) = P_n(S_1) + P_n(S_2). \tag{6.14.6}$$

Wegen $B \subset S_1 \subset A$ ist

$$P_n(B) \leqq P_n(S_1) \leqq P_n(A). \tag{6.14.7}$$

Wir bemerken weiter, daß

$$F_n\left(\frac{x}{a+\varepsilon}\right) = P\left(X_n < \frac{x}{a+\varepsilon}\right)$$

$$= P\left\{X_n < \frac{x}{a+\varepsilon},\ |Y_n - a| \leqq \varepsilon\right\} + P\left\{X_n < \frac{x}{a+\varepsilon},\ |Y_n - a| > \varepsilon\right\}$$

$$= P_n(B) + P\left\{X_n < \frac{x}{a+\varepsilon},\ |Y_n - a| > \varepsilon\right\},$$

$$F_n\left(\frac{x}{a-\varepsilon}\right) = P_n(A) + P\left\{X_n < \frac{x}{a-\varepsilon},\quad |Y_n - a| > \varepsilon\right\}$$

ist. Unter Berücksichtigung der Formeln (6.14.6) und (6.14.7) erhalten wir daraus die Ungleichung

$$F_n\left(\frac{x}{a+\varepsilon}\right) - P\left\{X_n < \frac{x}{a+\varepsilon},\quad |Y_n - a| > \varepsilon\right\}$$

$$\leqq \Phi_n(x) - P_n(S_2) \leqq F_n\left(\frac{x}{a-\varepsilon}\right) - P\left\{X_n < \frac{x}{a-\varepsilon},\quad |Y_n - a| > \varepsilon\right\}.$$

Es ist

$$P\left\{X_n < \frac{x}{a+\varepsilon},\quad |Y_n - a| > \varepsilon\right\} \leqq P\{|Y_n - a| > \varepsilon\},$$

$$P_n(S_2) = P\{X_n Y_n < x,\ |Y_n - a| > \varepsilon\} \leqq P\{|Y_n - a| > \varepsilon\},$$

$$P\left\{X_n < \frac{x}{a-\varepsilon},\quad |Y_n - a| > \varepsilon\right\} \leqq P\{|Y_n - a| > \varepsilon\}.$$

Aus der Voraussetzung über die stochastische Konvergenz der Folge $\{Y_n\}$ gegen die Konstante a folgt, daß für $n \to \infty$ die Wahrscheinlichkeiten auf der rechten Seite dieser Ungleichungen gegen Null streben. Wir erhalten also aus (6.14.8)

$$F\left(\frac{x}{a+\varepsilon}\right) = \lim_{n\to\infty} F_n\left(\frac{x}{a+\varepsilon}\right) \leqq \lim_{n\to\infty} \Phi_n(x) \leqq \lim_{n\to\infty} F_n\left(\frac{x}{a-\varepsilon}\right) = F\left(\frac{x}{a-\varepsilon}\right),$$

wobei $\dfrac{x}{a+\varepsilon}$ und $\dfrac{x}{a-\varepsilon}$ Stetigkeitsstellen der Verteilungsfunktion $F(x)$ sind. Beachtet man darüber hinaus, daß man ε beliebig klein wählen kann, so ergibt sich sofort (6.14.1).

Der hier gegebene Beweis des Satzes 6.14.1 geht auf CRAMÉR [2] zurück. Einen anderen Beweis für die Behauptung α) gab NEYMAN [8] (siehe auch WOLFOWITZ [1]). Einen einfachen Beweis aller Teile des Satzes 6.14.1 erhält man auch unter Verwendung des Satzes 6.4.1 (siehe SVERDRUP [1]).

Eine Verallgemeinerung der Behauptung α) des Satzes 6.14.1 für den Fall, daß die Folge der Verteilungsfunktionen der Zufallsvariablen Y_n gegen die Verteilungsfunktion einer nichtausgearteten Zufallsvariablen konvergiert, stammt von TEICHER [1].

B. Eine Folgerung aus diesem Satz ist der Satz von SLUTSKI [1]:

Satz 6.14.2. *Wenn die Folgen* $\{X_{1k}\}, \ldots, \{X_{rk}\}$ *von Zufallsvariablen, wobei* r *eine bestimmte Zahl und* $k = 1, 2, \ldots$ *ist, stochastisch gegen die festen Werte* $a_1, \ldots, a_r$ *konvergieren, dann strebt jede rationale Funktion* $R(X_{1k}, \ldots, X_{rk})$ *stochastisch gegen den festen Wert* $R(a_1, \ldots, a_r)$, *wenn dieser Wert endlich ist.*

Für den Beweis genügt es, in Satz 6.14.1 für $F(x)$ die Verteilungsfunktion der Einpunktverteilung zu nehmen und zu bemerken, daß dieser Satz richtig bleibt für eine Folge von Zufallsvariablen, die man erhält, indem man endlich viele rationale Operationen an den Zufallsvariablen ausführt.

C. Als einfache Folgerung von Satz 6.14.1 erhalten wir folgende Verallgemeinerung von Satz 6.2.1.

Satz 6.14.3. *Es sei* $\{Z_n\}$ $(n = 1, 2, 3, \ldots)$ *eine Folge von Zufallsvariablen, die stochastisch gegen die Zufallsvariable* Z_0 *konvergiert. Es besteht dann die Relation*

$$\lim_{n\to\infty} F_n(z) = F_0(z) \tag{6.14.9}$$

an jeder Stetigkeitsstelle der Verteilungsfunktion $F_0(z)$, *wobei* $F_n(z)$ *bzw.* $F_0(z)$ *die Verteilungsfunktionen der Zufallsvariablen* Z_n *bzw.* Z_0 *sind.*

Beweis. Es sei die Voraussetzung des Satzes erfüllt. Wir schreiben

$$Z_n = Z_0 + (Z_n - Z_0).$$

Da die Folge $\{Z_n - Z_0\}$ stochastisch gegen Null konvergiert, folgt die Formel (6.14.9) sofort aus der Behauptung α) des Satzes 6.14.1.

Es sei darauf hingewiesen, daß die Umkehrung dieses Satzes nicht gilt. Bekanntlich (siehe Satz 6.2.1) läßt sich dieser Satz in dem speziellen Fall umkehren, wenn Z_0 mit der Wahrscheinlichkeit 1 einen konstanten Wert annimmt.

6.15. Schlußbemerkungen

In den letzten Jahren erschienen Arbeiten, in denen Grenzverteilungen von Summen unabhängiger Zufallsvariabler behandelt werden, bei denen die Anzahl der Summanden eine Zufallsvariable ist. Es sei daran erinnert, daß wir in 5.14

eine Zufallsvariable Y betrachtet hatten, die durch die Formel

$$Y = \sum_{k=1}^{N} X_k$$

definiert ist, wobei nicht nur X_k, sondern auch N eine Zufallsvariable ist. Die Verteilung der Variablen N möge vom Parameter t abhängen. Dann ergibt sich das Problem, die Grenzverteilung der Zufallsvariablen Y (geeignet normiert) für den Fall zu bestimmen, daß $t \to \infty$ geht. Wir begnügen uns hier damit, einige der wichtigsten Arbeiten auf diesem Gebiet zu nennen, und zwar die Arbeiten von ROBBINS [1], DOBRUSCHIN [4] und ANSCOMBE [1]. Diese Arbeiten haben praktische Anwendungen gefunden (siehe z. B. RÉNYI [7] und TAKÁCS [5]). Auf eine ähnliche Problematik werden wir in Kapitel 17 stoßen (vgl. auch die Aufgaben 6.16.46 und 6.16.47).

6.16. Aufgaben und Ergänzungen

1. Die Zufallsvariablen X und Y seien unabhängig, und es habe X die Poissonsche Verteilung mit dem Parameter $\lambda = 1$.

   a) Man bestimme unter Verwendung von (6.5.6) die Verteilungsfunktion der Zufallsvariablen $Z = X + Y$, wobei Y der Binomialverteilung mit $n = 5$ und $p = \frac{1}{3}$ genügt.

   b) Man bestimme unter Verwendung von (6.5.9) die Verteilungsfunktion des Quotienten $Z = \frac{X}{Y}$, wobei Y eine Zufallsvariable mit der Normalverteilung $N(0; 1)$ ist.

2. Man beweise den nachstehenden *Satz von* HELLY: Konvergiert die Folge der Verteilungsfunktionen $F_n(x)$ $(n = 1, 2, \ldots)$ gegen die Verteilungsfunktion $F(x)$ und ist $g(x)$ eine auf der ganzen x-Achse stetige und beschränkte Funktion, so ist

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x)\, dF_n(x) = \int_{-\infty}^{\infty} g(x)\, dF(x).$$

   Hinweis. Der Beweis ist dem Beweis des Satzes 6.6.1a ähnlich.

3. Man gebe ein Beispiel für eine Folge von Verteilungsfunktionen $F_n(x)$ von Zufallsvariablen X_n mit den Varianzen $D^2(X_n)$, für die die Beziehung

$$\lim_{n\to\infty} F_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2}\, dx$$

   gilt, die Beziehung

$$\lim_{n\to\infty} D^2(X_n) = 1$$

   hingegen nicht erfüllt ist.

Hinweis. Man betrachte zunächst die Folge der Zufallsvariablen Y_n, für die

$$P(Y_n = 0) = 1 - \frac{1}{n}, \quad P(Y_n = n) = \frac{1}{n} \qquad (n = 1, 2, 3, \ldots)$$

gilt, und verwende dann die Behauptung α) des Satzes 6.14.1.

4. Man beweise die folgende Behauptung: Soll eine Folge von Verteilungsfunktionen $F_n(x)$ gegen die Verteilungsfunktion $F(x)$ konvergieren, so ist notwendig und hinreichend, daß die Beziehung

$$\lim_{n\to\infty} F_n(x) = F(x)$$

in einer gewissen Menge S erfüllt ist, die auf der ganzen x-Achse überall dicht ist.

5. Es möge die Folge der charakteristischen Funktionen $\varphi_n(t)$ gegen die charakteristische Funktion $\varphi(t)$, die Punktfolge $\{t_n\}$ gegen t_0 konvergieren. Man zeige, daß $\varphi_n(t_n) \to \varphi(t_0)$ gilt.

6. Man sagt, die Verteilungsfunktionen $F(x)$ und $G(x)$ seien *vom gleichen Typ*, wenn Konstanten $a > 0$ und b derart existieren, daß die Gleichung

$$G(x) = F(ax + b)$$

erfüllt ist. Die Folge der Verteilungsfunktionen $F_n(x)$ möge für $n \to \infty$ gegen die nicht ausgeartete Verteilungsfunktion $F(x)$ konvergieren. Ferner möge die Folge der Verteilungsfunktionen $F_n(a_n x + b_n)$ gegen eine nicht ausgeartete Verteilungsfunktion $G(x)$ konvergieren. Man zeige, daß dann $G(x)$ vom gleichen Typ ist wie $F(x)$ (Chintschin [8]).

Hinweis. Man beschränke sich auf Teilfolgen $\{a_{n_k}\}$ und $\{b_{n_k}\}$, die gegen die (endlichen oder unendlichen) Zahlen a bzw. b konvergieren. Sodann zeige man, daß $0 < a < \infty$ und $-\infty < b < \infty$ und daß $F_n(a_n x + b_n) \to F(ax + b)$ gilt.

(Eine Verallgemeinerung des hier verwendeten Begriffs „Typ" sowie des oben angegebenen Satzes für n-dimensionale Zufallsvektoren stammt von Fisz [2].)

7. Man beweise die Beziehung (6.4.6).

Hinweis. Der Rand der Menge S ist $\mathrm{Fr}(S) = \overline{S} \cap \overline{R - S}$; die Menge der inneren Punkte von S ist $\mathrm{Int}(S) = S - \mathrm{Fr}(S)$. Man verwende die Gleichung $P_n[\mathrm{Int}(S)] + P_n[\mathrm{Int}(R - S)] + P_n[\mathrm{Fr}(S)] = 1$ $(n = 1, 2, \ldots)$ und zeige zuerst, daß aus (6.4.1) sowie aus der Gleichung $P[\mathrm{Fr}(S)] = 0$ die Beziehungen $\lim_{n\to\infty} P_n[\mathrm{Int}(S)] = P[\mathrm{Int}(S)]$, $\lim_{n\to\infty} P_n[\mathrm{Int}(R - S)] = P[\mathrm{Int}(R - S)]$ und $\lim_{n\to\infty} P_n[\mathrm{Fr}(S)] = P[\mathrm{Fr}(S)]$ folgen.

8. (Umkehrung des Satzes von Lindeberg-Lévy). Es sei $\{X_k\}$ $(k = 1, 2, 3, \ldots)$ eine Folge unabhängiger Zufallsvariabler mit gleicher Verteilung. Ferner möge für gewisse Konstanten a und A_n $(n = 1, 2, 3, \ldots)$ die Beziehung

$$\lim_{n\to\infty} P\left(\frac{1}{a\sqrt{n}} \sum_{k=1}^{n} X_k - A_n < z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{z^2}{2}\right) dz$$

für jedes z gelten. Man zeige, daß die Varianz σ^2 der Zufallsvariablen X_k existiert. Es gilt dann $a = \sigma$, und für A_n läßt sich die Größe $\frac{\sqrt{n}}{\sigma} E(X_k)$ verwenden (vgl. GNEDENKO und KOLMOGOROV [1], § 35).

9. a) Die Zufallsvariablen X_i ($i = 1, 2, \ldots, r$) seien unabhängig, und es mögen die X_i der Poissonschen Verteilung mit dem Parameter $\lambda_i > 0$ genügen. Man zeige, daß $\sum_{i=1}^{r} X_i$ für $\lambda_1 + \cdots + \lambda_r \to \infty$ die asymptotische Normalverteilung $N\left(\sum_{i=1}^{r} \lambda_i; \sqrt{\sum_{i=1}^{r} \lambda_i}\right)$ hat.

   b) Die Zufallsvariablen X_1 bzw. X_2 mögen eine Poissonsche Verteilung mit den Parametern $\lambda_1 > 0$ bzw. $\lambda_2 > 0$ haben. Man zeige, daß für $\lim\limits_{\substack{\lambda_1 \to \infty \\ \lambda_2 \to \infty}} (\lambda_1/\lambda_2) = 1$ die Zufallsvariable

$$Y = \frac{X_1 - X_2}{(X_1 + X_2)^p}$$

   für ein beliebiges $p > 0$ für $\lambda_1 \to \infty$ und $\lambda_2 \to \infty$ die asymptotische Normalverteilung

$$N\left(\frac{\lambda_1 - \lambda_2}{(\lambda_1 + \lambda_2)^p}; \frac{1}{(\lambda_1 + \lambda_2)^{p-0,5}}\right)$$

   besitzt.

10. Die Zufallsvariablen X_1, X_2 und X mögen die gleiche nicht ausgeartete Verteilung mit der endlichen Varianz σ^2 besitzen, und es seien X_1 und X_2 unabhängig. Gilt dann für ein gewisses $a > 0$ die Gleichung

$$X_1 + X_2 = aX,$$

so haben X_1, X_2 und X die Normalverteilung $N(0; \sigma)$, und es ist $a = \sqrt{2}$ (PÓLYA [2]).
Hinweis. Nach Verifizierung der Tatsache, daß $a = \sqrt{2}$ ist und die charakteristische Funktion $\varphi(t)$ der Zufallsvariablen X für jedes natürliche k der Gleichung

$$\varphi^{2^k}\left(t/\sqrt{2^k}\right) = \varphi(t)$$

genügt, wende man den Satz von LINDEBERG-LÉVY an.
(Eine Verallgemeinerung für mehrdimensionale Zufallsvariable und Zufallsvektoren in einem Hilbertraum haben PROCHOROW und FISZ [1] angegeben.)

11. Man zeige, daß unter den Voraussetzungen des Satzes 6.9.1 auch die Bedingung (6.9.20) erfüllt ist.

12. Es sei $\{X_k\}$ ($k = 1, 2, \ldots$) eine Folge unabhängiger Zufallsvariabler, und es sei

$$P(X_k = -1) = P(X_k = 1) = \frac{1 - 2^{-k}}{2},$$

$$P(X_k = -2^k) = P(X_k = 2^k) = \frac{1}{2^{k+1}}.$$

Man überprüfe, ob für diese Folge die Bedingung (6.9.20) erfüllt ist.

13. a) Man überprüfe, ob für die Folgen $\{X_i\}$ und $\{Y_i\}$ unabhängiger Zufallsvariabler, deren

Wahrscheinlichkeitsfunktionen durch die Formeln (6.12.25) und (6.12.26) gegeben sind, die Bedingung (6.9.20) erfüllt ist.

b) Man löse die gleiche Aufgabe für den Fall, daß $P(X_k = -k^a) = P(X_k = k^a) = \frac{1}{2}$ für $a > 0$ gilt.

14. Die Zufallsvariablen X_k $(k = 1, 2, \ldots)$ seien unabhängig und mögen die Normalverteilung $N\left(0; 2^{-\frac{k}{2}}\right)$ besitzen. Man überprüfe, ob für diese Folge a) der zentrale Grenzwertsatz gilt, b) die Bedingung (6.9.20) erfüllt ist, c) die Beziehung $\lim_{n\to\infty} \max_{1\leq k\leq n} (\sigma_k/C_n) = 0$ gilt.

15. Wir betrachten die Folge unabhängiger Zufallsvariabler X_k $(k = 0, 1, 2, \ldots)$, wobei X_0 die durch (5.8.6) definierte Gammaverteilung hat, während die X_k $(k = 1, 2, \ldots)$ die gleichen Verteilungen haben wie in der vorhergehenden Aufgabe. Gilt für die Folge $\{X_k\}$ $(k = 0, 1, 2, \ldots)$ der zentrale Grenzwertsatz?

16. Die Zufallsvariablen X_k $(k = 1, 2, \ldots)$ seien unabhängig und mögen die Poissonsche Verteilung mit dem Parameter $\lambda_k = 2^{-k}$ besitzen. Man stelle fest, ob für die Folge $\{X_k\}$ der zentrale Grenzwertsatz gilt.

17. Es sei $Y_n = \sum_{k=1}^{k_n} X_{nk}$ (vgl. 6.9.C), wobei die Zufallsvariablen X_{nk} $(k = 1, 2, \ldots, k_n)$ für jedes n unabhängig sind und die durch die Formeln $P(X_{nk} = x_{nkl}) = p_{nkl}$ gegebenen Wahrscheinlichkeitsfunktionen haben; dabei sei $\sum_{l=1}^{r} P_{nkl} = 1$ $(n = 1, 2, \ldots;\ k = 1, 2, \ldots, k_n;\ l = 1, 2, \ldots, r;\ r > 2)$. Wir nehmen an, daß

a) die X_{nk} *asymptotisch konstant* sind, d. h., daß für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} \max_{1\leq k\leq k_n} P(|X_{nk} - m_{nk}| > \varepsilon) = 0$$

gilt, wobei m_{nk} der Mittelwert der Variablen X_{nk} ist;

b) $\lim_{n\to\infty} \max_{1\leq k\leq k_n} z_{nkl} = \lim_{n\to\infty} \min_{1\leq k\leq k_n} z_{nkl} = z_l$ $(l = 1, \ldots, r)$ mit $z_{nkl} = x_{nk,l+1} - x_{nkl}$ gilt.

Man bestimme die Klassen der möglichen Grenzverteilungsfunktionen der Folge der Verteilungsfunktionen $F_n(y)$ der Zufallsvariablen $Y_n - A_n$ für beliebige Konstantenfolgen $\{A_n\}$ (Kubik [1]).

18. Wir bezeichnen mit $m_k^{(n)}$ das Moment der Ordnung k der Zufallsvariablen X_n mit der Verteilungsfunktion $F_n(x)$. Für $k = 1, 2, \ldots$ mögen die endlichen Grenzwerte $m_k = \lim_{n\to\infty} m_k^{(n)}$ existieren, und diese Grenzwerte mögen darüber hinaus (vgl. Aufgabe 4.8.10) eine gewisse Verteilungsfunktion $F(x)$ eindeutig bestimmen. Man zeige, daß dann die Folge $\{F_n(x)\}$ gegen $F(x)$ konvergiert. (Das ist der *zweite zentrale Grenzwertsatz*; vgl. Fréchet und Shohat [1].)

19. Die Zufallsvariablen $X_1, X_2, X_3, \ldots$ mögen allen Bedingungen des Satzes von Lindeberg-Lévy genügen, und es möge darüber hinaus $E(|X_i|^3)$ existieren. Dann gilt die Beziehung

$$|F_n(z) - \Phi(z)| \leqq \frac{E(|X_i|^3)}{\sigma^3} \frac{1}{\sqrt{n}},$$

wobei c eine Konstante ist (Cramér [1], Esseen [1], Berry [1]).

20. Die Zufallsvariablen X_i $(i = 1, 2, \ldots)$ seien unabhängig und mögen die gleiche durch die Formeln

$$P(X_i = 0) = P(X_i = 3) = P(X_i = 7) = P(X_i = 12) = \frac{1}{4}$$

gegebene Verteilung besitzen. Man stelle fest, ob für die Folge $\{X_i\}$ der lokale Grenzwertsatz von GNEDENKO gilt.

21. Die Zufallsvariable X möge eine Poissonsche Verteilung mit dem Parameter λ haben. Ferner sei $u_r = P(X = r)$ $(r = 0, 1, 2, \ldots)$, $t = (r - \lambda)/\sqrt{\lambda}$ und

$$v_r = \frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{t^2}{2}\right),$$

$$w_r = v_r\left(1 - \frac{t}{2\sqrt{\lambda}} - \frac{t^3}{6\sqrt{\lambda}}\right).$$

Des weiteren gelte $\lambda \to \infty$ und $r \to \infty$ in der Weise, daß t absolut beschränkt ist; man zeige unter Anwendung der Stirlingschen Formel, daß für jedes $\varepsilon > 0$

$$\lim_{\lambda\to\infty} \left(\lambda^{1-\varepsilon}(u_r - v_r)\right) = 0,$$

$$\lim_{\lambda\to\infty} \left(\lambda^{\frac{3}{2}-\varepsilon}(u_r - w_r)\right) = 0$$

gilt.

22. Die Zufallsvariablen X_i $(i = 1, 2, \ldots)$ mögen allen Bedingungen des Satzes 6.10.1 und der Bedingung (w) genügen und das Moment der Ordnung $v \geqq 3$ besitzen. Dann gilt (mit den Bezeichnungen von 6.10) für $n \to \infty$ die asymptotische Formel

$$P_n(k) = \frac{1}{\sigma\sqrt{n}}\left(\varphi(z) + \sum_{r=1}^{v-2} n^{-r/2} U_r\left(-\varphi(z)\right)\right) + o(n^{-(v-1)/2}),$$

wobei $z = z_{nk}$, $\varphi(z) = \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{z^2}{2}\right)$ ist und die Funktionen U_r für jedes r sich durch Koeffizientenvergleich bei den aufeinanderfolgenden Potenzen von $1/\sqrt{n}$ in der Identität

$$\exp\left(\frac{\varkappa_{r+2}}{(r+2)!}\left(-\varphi(z)\right)^{(r+2)} n^{-r/2}\right) = 1 + \sum_{r=1}^{\infty} U_r\left(-\varphi(z)\right) n^{-r/2}$$

bestimmen lassen, in der $\varkappa_{r+2}$ die Semiinvariante der Ordnung $r + 2$ der Zufallsvariablen X_i/σ ist (ESSEEN [1]).

a) Man drücke $\left(-\varphi(z)\right)^{(r+2)}$ mit Hilfe Hermitescher Polynome aus.

b) Man berechne den Ausdruck $P_n(k)$ für $v = 4$.

23. Die unabhängigen Zufallsvariablen X_i $(i = 1, 2)$ mögen eine Poissonsche Verteilung mit den Parametern $\lambda_1 > 0$ bzw. $\lambda_2 > 0$ haben. Dabei sei $\lambda_i = d_i n$, wobei d_1 und d_2 positive Konstanten sind.

a) Man beweise: Ist $\lambda_i \to \infty$ $(i = 1, 2)$, so gilt für eine beliebige ganze Zahl $u \geqq 1$ die asymptotische Formel

$$P(X_2 - X_1 = k) = \frac{1}{\sqrt{\lambda_1 + \lambda_2}} \left(\varphi(z) + \sum_{r=1}^{u} U_r\left(-\varphi(z)\right) \right) + o\left((\lambda_1 + \lambda_2)^{-(u+1)/2} \right),$$

wobei $z = \dfrac{k - (\lambda_1 + \lambda_2)}{\sqrt{\lambda_1 + \lambda_2}}$ ist und $U_r\left(-\varphi(z)\right)$ ebenso bestimmt wird wie in Aufgabe 22, und zwar mit

$$\varkappa_{r+2} = \begin{cases} n^{r/2} (\lambda_1 + \lambda_2)^{-r/2} & \text{für} \quad r = 2, 4, \ldots, \\ n^{r/2} (\lambda_2 - \lambda_1)(\lambda_1 + \lambda_2)^{-(r+2)/2} & \text{für} \quad r = 1, 3, \ldots \end{cases}$$

b) Unter Verwendung der Aufgabe 5.14.5b ermittle man die asymptotische Formel für die Besselsche Funktion $I_k(2\lambda)$ mit rein imaginärem Argument.

24. Unter Verwendung des in Aufgabe 22 angegebenen Ergebnisses von Esseen bestimme man für u_r eine genauere asymptotische Formel als die in Aufgabe 21 hergeleitete.

25. Die Zufallsvariablen $X_1, X_2, \ldots$ seien unabhängig und mögen die gleiche Verteilung mit der durch die Formel

$$f(x) = \begin{cases} 0 & \text{für} \quad |x| \geqq \dfrac{1}{e}, \\ \dfrac{1}{2|x| \log^2 |x|} & \text{für} \quad |x| < \dfrac{1}{e} \end{cases}$$

definierten Dichtefunktion $f(x)$ besitzen. Man zeige, daß der integrale Grenzwertsatz von Lindeberg-Lévy für die Folge $\{X_i\}$ erfüllt ist, während die Folge der Dichtefunktionen der Zufallsvariablen $(X_1 + \cdots + X_n)/\sigma\sqrt{n}$ nicht gegen die Dichtefunktion der Normalverteilung konvergiert (Gnedenko und Kolmogorov [1]).

26. Es sei $\{X_i\}$ $(i = 1, 2, \ldots)$ eine Folge unabhängiger Zufallsvariabler mit gleicher Dichtefunktion $f(x)$, und es sei $f_n(x)$ die Dichte der Zufallsvariablen

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{B_n} - A_n.$$

Sollen dann die Konstantenfolgen $\{A_n\}$ und $\{B_n\}$ existieren, für die die Beziehung

$$\lim_{n\to\infty} \sup_x \left(f_n(x) - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x^2 \right) \right) = 0 \tag{*}$$

gilt, so ist notwendig und hinreichend, daß die folgenden Bedingungen erfüllt sind:

1. Die Folge der Verteilungsfunktionen $F_n(x)$ der Variablen Z_n konvergiert gegen $\Phi(x)$;

2. es existiert ein m derart, daß die Verteilungsfunktion der Summe $X_1 + X_2 + \cdots + X_m$ absolut stetig ist und eine beschränkte Ableitung besitzt (Gnedenko [2], [3]).

a) Man stelle fest, ob für eine Folge unabhängiger Zufallsvariabler X_i mit der gleichen Rechtecksverteilung im Intervall $[0, 1]$ die Beziehung (*) gilt.

b) Man löse die gleiche Aufgabe für den Fall, daß die X_i eine Gammaverteilung haben.

c) Man löse die gleiche Aufgabe für den Fall, daß die X_i eine Betaverteilung haben.

27. Man beweise, daß der Satz 6.11.3 gültig bleibt, wenn

a) die Voraussetzung von der Gleichheit der Varianz durch die Annahme ersetzt wird, daß die Varianzen gleichmäßig beschränkt sind;

b) die Voraussetzung, daß die Zufallsvariablen X_k paarweise nicht korreliert sind, durch die Annahme ersetzt wird, daß die Kovarianzen der Paare (X_{k_1}, X_{k_2}) für $|k_1 - k_2| \to \infty$ gleichmäßig gegen Null konvergieren.

28. Die Zufallsvariablen $X_1, X_2, X_3, \ldots$ und X seien gleichmäßig beschränkt. Man zeige, daß die Folge $\{X_k\}$ dann und nur dann stochastisch gegen Null konvergiert, wenn

$$\lim_{n\to\infty} E(|X_n - X|^2) = 0$$

gilt.

Hinweis. Man verwende die Tschebyscheffsche Ungleichung und die Kolmogoroffsche Ungleichung aus Aufgabe 3.9.8.

29. Es sei $\{A_k\}$ $(k = 1, 2, \ldots)$ eine Folge zufälliger (abhängiger oder unabhängiger) Ereignisse. Wir betrachten die Folge der Zufallsvariablen Y_k mit $P(Y_k = 1) = P(A_k)$, $P(Y_k = 0) = 1 - P(A_k)$. Dabei setzen wir

$$X_n = \frac{1}{n} \sum_{k=1}^{n} \left(Y_k - P(A_k)\right),$$

$$p_1(n) = \frac{1}{n} \sum_{k=1}^{n} P(A_k),$$

$$p_2(n) = \frac{2}{n(n-1)} \sum_{1<j<k<n} P(A_j A_k),$$

$$d_n = p_2(n) - p_1^2(n).$$

Man zeige, daß die Folge $\{X_n\}$ dann und nur dann stochastisch gegen Null konvergiert, wenn $\lim_{n\to\infty} d_n = 0$ ist. Man formuliere dieses Ergebnis mit Hilfe der Häufigkeit der erfolgreichen Versuche unter n Versuchen insgesamt.

30. (Mit Bezeichnungen von Aufgabe 29.) Man zeige: Gilt

$$\sum_{n=1}^{\infty} \frac{d_n}{n} < \infty,$$

so konvergiert (Dvoretzky [1]) die Folge $\{X_n\}$ mit der Wahrscheinlichkeit 1 gegen Null.

Hinweis. Man beweise zunächst den folgenden Hilfssatz: Es sei $a_n \geqq 0$ $(n = 1, 2, \ldots)$, ferner

$$\sum_{n=1}^{\infty} \frac{a_n}{n} < \infty.$$

Es existiert dann eine Folge $\{n_i\}$ natürlicher Zahlen derart, daß $n_{i+1} - n_i = o(n_i)$ und $\sum_{i=1}^{\infty} a_{n_i} < \infty$ für $n_i \to \infty$ ist.

31. Aus der vorhergehenden Aufgabe ziehe man die Folgerung, daß die Häufigkeit der erfolgreichen Versuche in der Folge der paarweise unabhängigen Ereignisse A_k, für die $P(A_k) = p = \text{const}$ ist, mit der Wahrscheinlichkeit 1 gegen p konvergiert. (Man vergleiche hierzu auch Aufgabe 40.)

32. Die Ereignisse A_k $(k = 1, 2, \ldots)$ mögen die gleiche Wahrscheinlichkeit $P(A_k) = p$ $(0 < p < 1)$ haben.

    a) Man zeige, daß $P(\limsup_{k\to\infty} A_k) \geqq p$ ist.

    b) Man gebe ein Beispiel einer Folge $\{A_k\}$ an, für die $P(\limsup_{k\to\infty} A_k) = p$ gilt.

33. Wir betrachten die Folge der Ereignisse A_k $(k = 1, 2, \ldots)$ mit den Wahrscheinlichkeiten $P(A_k)$ und die Zufallsvariablen Y_k mit

$$P(Y_k = 1) = P(A_k), \qquad P(Y_k = 0) = 1 - P(A_k).$$

Wir nehmen an, daß $0 < P(A_1) < 1$ ist und daß für $k = 1, 2, \ldots$ und für beliebige Werte $y_1, y_2, \ldots, y_k$ der Zufallsvariablen $Y_1, Y_2, \ldots, Y_k$ die Beziehungen

$$0 < p_k' \leqq P(A_{k+1} | y_1, \ldots, y_k) \leqq p_k'' < 1,$$
$$p_k' \leqq P(A_k) \leqq p_k''$$

gelten. Man zeige, daß dann die folgenden Relationen erfüllt sind (Borel [1], [2]):

$$\text{Ist } \sum_k p_k'' < \infty, \text{ so ist } P(\limsup_{k\to\infty} A_k) = 0;$$
$$\text{ist } \sum_k p_k' = \infty, \text{ so ist } P(\limsup_{k\to\infty} A_k) = 1.$$

34. (Fortsetzung).

    a) Wir setzen für $k > n$

$$p_{nk} = P(A_{k+1} | y_n = y_{n+1} = \cdots = y_k = 0),$$
$$p_{nn} = P(A_n).$$

Man zeige, daß $P(\limsup_{k\to\infty} A_k) = 0$ dann und nur dann gilt, wenn $\lim_{n\to\infty} \sum_{k=n}^{\infty} p_{nk} = 0$ ist (Loève [3]).

    b) Man zeige, daß $P(\limsup_{k\to\infty} A_k) = 1$ dann und nur dann gilt, wenn $\sum_{k=1}^{\infty} P(A_{k+1} | y_1, \ldots, y_k) = \infty$ ist, wobei nur die Folgen $y_1, y_2, y_3, \ldots$ zu berücksichtigen sind, in denen endlich oft $y_k = 1$ und $P(y_1, y_2, y_3, \ldots, y_k) \neq 0$ ist (Nash [1]).

35. (Verallgemeinerung des Lemmas von Borel-Cantelli, die wir Erdös und Rényi [2] verdanken.) Es sei $\{A_k\}$ $(k = 1, 2, \ldots)$ eine Folge zufälliger Ereignisse derart, daß

$$\sum_{k=1}^{\infty} P(A_k) = \infty \qquad (*)$$

gilt.

a) Man zeige, daß $P(\limsup_{k\to\infty} A_k) = 1$ gilt, wenn neben (*) die Beziehung

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \sum_{l=1}^{n} P(A_k A_l)}{\left[\sum_{k=1}^{n} P(A_k)\right]^2} = 1 \qquad (**)$$

erfüllt ist.

b) Man beweise unter Verwendung des Ergebnisses von a), daß die Behauptung von a) gültig bleibt, falls die Formel (*) gilt und die Formel (**) entweder durch die Annahme ersetzt wird, daß die Ereignisse A_k paarweise unabhängig sind, oder aber durch die Voraussetzung, daß für alle k und l ($k \neq l$) die Beziehung

$$P(A_k A_l) \leqq P(A_k)\, P(A_l)$$

gilt.

36. Wir setzen $p_k = P(A_k)$ und $v_k = P(A_{k+1} | \bar{A}_1, \ldots, \bar{A}_k)$ mit $0 < p_k < 1$, $0 < v_k < 1$ ($k = 1, 2, \ldots$).

a) Man zeige: Eine notwendige und hinreichende Bedingung dafür, daß

$$P\left(\bigcap_{k=1}^{\infty} \bar{A}_k\right) > 0$$

gilt, ist

$$\sum_k v_k = \infty. \qquad (*)$$

b) Man zeige: Gilt die Gleichung (*) und darüber hinaus

$$\sum_k p_k = \infty, \qquad (**)$$

so ist $P(\limsup_{k\to\infty} A_k) = 1$.

c) Man zeige, daß die Formel (**) nicht aus (*) folgt.

37. Kann der zentrale Grenzwertsatz für eine Folge unabhängiger Zufallsvariabler gelten, obwohl das Gesetz der großen Zahlen für diese Folge nicht gilt? $\left(\text{Man vergleiche hierzu Aufgabe 13b für } a \geqq \frac{1}{2}.\right)$ Wie lautet die Antwort, wenn wir annehmen, daß die Zufallsvariablen der betrachteten Folge die gleiche Verteilung haben?

38. a) Die Zufallsvariablen X_k ($k = 1, 2, \ldots$) seien unabhängig und mögen die gleiche Verteilungsfunktion $F(x)$ haben. Man zeige, daß die Folge

$$\left\{\frac{X_1 + \cdots + X_n}{n}\right\}$$

dann und nur dann gegen eine gewisse Konstante a stochastisch konvergiert, wenn die charakteristische Funktion $\varphi(t)$ der Zufallsvariablen X_k an der Stelle 0 differenzierbar ist und $\varphi'(0) = ia$ gilt.

Hinweis. *Verfahren 1.* Man verwende den Satz von KOLMOGOROFF [7], nach dem, wenn die Folge $\{Y_n\}$ stochastisch gegen a konvergiert, die Beziehungen $\lim_{x\to\infty} xP(|X_k| > x) = 0$ und $a = \lim_{x\to\infty} \int_{-x}^{x} x\,dF(x)$ gelten, und stütze sich anschließend auf das Ergebnis von PITMAN (siehe Aufgabe 4.8.9b).

Verfahren 2 (EHRENFEUCHT und FISZ [1]). Man verwende den Hilfssatz: Ist die Funktion $\psi(t)$ in den Intervallen $(-\infty, 0)$ und $(0, \infty)$ stetig und gilt $\lim_{n\to\infty} \psi\left(\frac{t}{n}\right) = 0$ für jedes $t \neq 0$, so ist $\lim_{t\to 0} \psi(t) = 0$.

Siehe auch DUGUÉ [2].

b) Man konstruiere eine Folge unabhängiger Zufallsvariabler, für die das schwache Gesetz der großen Zahlen im Sinne der Aufgabe 38a gilt, während das starke Gesetz der großen Zahlen nicht gilt.

39. Es sei $\{X_k\}$ $(k = 1, 2, \ldots)$ eine Folge unabhängiger Zufallsvariabler, für die $E(X_k) = 0$ ist $(k = 1, 2, \ldots)$. Wir definieren die folgenden Zufallsvariablen:

$$X_{nk} = \begin{cases} X_k & \text{für } |X_k| < n, \\ 0 & \text{für } |X_k| \geqq n, \end{cases} \qquad Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Man zeige, daß (KOLMOGOROFF [1]) die Folge $\{Y_n\}$ dann und nur dann stochastisch gegen Null konvergiert, wenn die drei nachstehenden Beziehungen erfüllt sind:

$$\sum_{k=1}^{n} P(|X_k| \geqq n) \to 0, \tag{*}$$

$$\frac{1}{n}\sum_{k=1}^{n} E(X_{nk}) \to 0, \tag{**}$$

$$\frac{1}{n^2}\sum_{k=1}^{n} D^2(X_{nk}) \to 0. \tag{***}$$

Anmerkung. Wie BREIMAN [1] gezeigt hat, läßt sich die Beziehung (***) nicht durch

$$\frac{1}{n^2}\sum_{k=1}^{n} E(X_{nk}^2) \to 0$$

ersetzen.

40. Es sei $\{X_k\}$ eine Folge paarweise unabhängiger Zufallsvariabler. Man zeige: Existieren die Varianzen $D^2(X_k)$ $(k = 1, 2, \ldots)$ und sind diese gleichmäßig beschränkt, so gilt für die Folge $\{X_k\}$ das starke Gesetz der großen Zahlen (RAJCHMAN [1]).

41. Es seien die X_i $(i = 1, 2, \ldots, n)$ unabhängige Zufallsvariable, und es sei $E(X_i) = 0$ und

$0 < D^2(X_i)$ $(i = 1, 2, \ldots, n)$. Ferner sei $\{c_i\}$ eine nichtwachsende Folge positiver Konstanten. Wir setzen

$$Y_k = \sum_{i=1}^{k} X_i \qquad (k = 1, 2, \ldots, n).$$

a) Man zeige, daß für jedes $\varepsilon > 0$ sowie für alle j und l $(1 \leqq j < l \leqq n)$ die folgende Ungleichung gilt (HÁJEK und RÉNYI [1]):

$$P\left(\max_{j \leqq k \leqq l} c_k |Y_k| \geqq \varepsilon\right) \leqq \frac{1}{\varepsilon^2}\left(c_j^2 D^2(Y_j) + \sum_{i=j+1}^{l} c_i^2 D^2(X_i)\right).$$

b) Man leite unter Verwendung dieser Ungleichung den Satz 6.12.1 her (der Beweis dieses Satzes wird sich dadurch erheblich vereinfachen).

In der nachfolgenden Aufgabe führen wir das *Gesetz des iterierten Logarithmus* für die Binomialverteilung an, das von CHINTSCHIN [8] bewiesen wurde. Verallgemeinerungen stammen von KOLMOGOROFF [16], ERDÖS [1] und FELLER [10].

42. Die Zufallsvariablen X_n $(n = 1, 2, \ldots)$ mögen eine Binomialverteilung haben, und es sei Y_n durch (6.7.2) definiert. Man zeige, daß dann

$$P\left(\limsup_{n\to\infty} \frac{Y_n}{\sqrt{2 \log \log n}} = 1\right) = 1,$$

$$P\left(\liminf_{n\to\infty} \frac{Y_n}{\sqrt{2 \log \log n}} = -1\right) = 1$$

gilt.

Hinweis. Man verwende die Ungleichung von Aufgabe 41.

43. Es sei $\{X_k\}$ $(k = 1, 2, \ldots)$ eine Folge unabhängiger Zufallsvariabler. Man zeige, daß dann

a) die Wahrscheinlichkeit für die Konvergenz der Reihe

$$\sum_{k=1}^{\infty} X_k = Y \tag{*}$$

entweder 1 oder 0 ist;

b) die Reihe (*) dann und nur dann stochastisch konvergiert, wenn sie mit der Wahrscheinlichkeit 1 konvergent ist (LÉVY [3], MARCINKIEWICZ und ZYGMUND [1]);

c) wenn die beiden Reihen

$$\sum_{k=1}^{\infty} E(X_k), \quad \sum_{k=1}^{\infty} D^2(X_k)$$

konvergent sind, die Reihe (*) mit der Wahrscheinlichkeit 1 konvergiert und darüber hinaus

$$E(Y) = \sum_{k=1}^{\infty} E(X_k),$$

$$D^2(Y) = \sum_{k=1}^{\infty} D^2(X_k)$$

ist.

44. Es sei $\{X_k\}$ $(k = 1, 2, \ldots)$ eine Folge unabhängiger Zufallsvariabler. Wir setzen

$$Z_k = \begin{cases} X_k & \text{für } |X_k| \leqq 1, \\ 0 & \text{für } |X_k| > 1. \end{cases}$$

Man zeige, daß unter dieser Bedingung die Reihe (*) der Aufgabe 43 dann und nur dann mit der Wahrscheinlichkeit 1 konvergiert, wenn gleichzeitig die nachstehenden drei Reihen konvergent sind:

$$\sum_{k=1}^{\infty} P(|X_k| > 1), \quad \sum_{k=1}^{\infty} E(Z_k), \quad \sum_{k=1}^{\infty} D^2(Z_k).$$

Man nennt diesen Satz den *Dreireihensatz* (Chintschin und Kolmogoroff [1]).

45. Der Vektor $(X_1, X_2, \ldots, X_r)$ möge die durch (6.13.1) gegebene Verteilung haben; hierbei können die Wahrscheinlichkeiten p_j Funktionen von n sein, weshalb wir sie durch p_{nj} bezeichnen. Man bestimme die Klasse der möglichen Grenzverteilungen der Folge der Verteilungsfunktionen der Zufallsvektoren $(A_{n1}X_1 + B_{n1}, \ldots, A_{nr}X_r + B_{nr})$ für beliebige Konstantenfolgen $\{A_{nj}\}$ und $\{B_{nj}\}$ (Fisz [2]).

Hinweis. Man verwende die mehrdimensionalen Analoga des Satzes 6.6.1 und den Satz von Chintschin (vgl. Aufgabe 6).

46. Es sei $\{X_k\}$ eine Folge unabhängiger Zufallsvariabler mit gleichen Verteilungen, und es existiere $m = E(X_k)$ und $\sigma^2 = D^2(X_k)$. Ferner sei N eine ganzzahlige, von $X_1, X_2, \ldots$ unabhängige Zufallsvariable mit dem Erwartungswert $c = E(N)$ und mit $d^2 = D^2(N)$. Ferner möge die Verteilung von N vom Parameter λ abhängen. Wir setzen $X = \sum_{k=1}^{n} X_k$.

a) Man zeige, daß $E(X) = mc$ und $D^2(X) = c\sigma^2 + m^2 d^2$ ist.

b) Man beweise: Hat N für $\lambda \to \infty$ die asymptotische Normalverteilung $N(c; d)$, so besitzt X die asymptotische Normalverteilung $N\left(mc; \sqrt{c\sigma^2 + m^2 d^2}\right)$. (Vergleiche Robbins [1]; eine weitgehende Verallgemeinerung stammt von Dobruschin [4].)

c) Man bestimme die Grenzverteilung der Zufallsvariablen X mit zusammengesetzter Binomialverteilung (siehe 5.13.A) für $\lambda \to \infty$.

47. Die Folge $\{X_k\}$ $(k = 1, 2, \ldots)$ möge alle Bedingungen des Satzes 6.8.1 für $m = 0$ und $\sigma = 1$ erfüllen, und es sei N eine ganzzahlige Zufallsvariable, deren Verteilung von $\lambda > 0$ abhängig ist; ferner möge $\dfrac{N}{\lambda}$ für $\lambda \to \infty$ stochastisch gegen eine positive Konstante konvergieren. Man zeige, daß dann die Zufallsvariable $\dfrac{X}{\sqrt{N}}$ mit $X = \sum_{k=1}^{N} X_k$ die Grenzverteilung $N(0; 1)$ besitzt (Anscombe [1]; siehe auch Rényi [7]).

48. Man beweise: Konvergiert eine Folge unbegrenzt teilbarer Verteilungsfunktionen (vgl. Aufgabe 5.14.25) gegen eine nicht ausgeartete Verteilungsfunktion $F(x)$, so ist auch $F(x)$ eine unbegrenzt teilbare Verteilungsfunktion.

49. Man beweise (Pólya [3]): Konvergiert die Folge der Verteilungsfunktionen $F_n(x)$ für $n \to \infty$ gegen eine stetige Verteilungsfunktion $F(x)$, so konvergiert sie gegen $F(x)$ auch gleichmäßig bezüglich x.
Wegen einer Verallgemeinerung dieses Ergebnisses siehe Ranga-Rao [1].

50. Der Umfang einer Kreisscheibe, die um eine drehbare Achse an einem Tisch befestigt ist, sei in $2m$ Bogenstücke unterteilt, die abwechselnd rot bzw. schwarz gekennzeichnet sind und die Zentriwinkel δ_1 bzw. δ_2 haben. Anfänglich markiert ein am Tisch befestigter Pfeil einen Nullpunkt auf dem Umfang der Kreisscheibe. Nun wird die Kreisscheibe in eine Drehbewegung versetzt. Wir bezeichnen mit A das zufällige Ereignis, das darin besteht, daß der Pfeil nach Stehenbleiben der Kreisscheibe auf einen roten Punkt auf dem Umfang der Kreisscheibe zeigt. Ferner bezeichnen wir mit X den Winkel, um den sich die Kreisscheibe dreht. Ist $F(x)$ die Verteilungsfunktion der Zufallsvariablen X, so gilt

$$P(A) = \int_{-\infty}^{\infty} h(x)\, dF(x),$$

wobei $h(x) = 1$ für $k(\delta_1 + \delta_2) \leqq x \leqq k(\delta_1 + \delta_2) + \delta_1$ $(k = 0, 1, 2, \ldots)$ und $h(x) = 0$ für alle übrigen x ist.

a) Man zeige: Hat X eine beliebige Dichtefunktion $f(x)$, so ist $\lim P(A) = \delta_1/(\delta_1 + \delta_2)$, wenn $m \to \infty$, $\delta_1 \to 0$ und $\delta_2 \to 0$ geht, wobei δ_1/δ_2 konstant bleibt und $f(x)$ fest gehalten wird (Poincaré [1], Fréchet [1], Hostinsky [1]).

b) Nun seien δ_1 und δ_2 konstant. Wir betrachten aufeinanderfolgende Umdrehungen der Kreisscheibe, wobei wir bei den einzelnen Umdrehungen die Kreisscheibe nicht in die Ausgangslage zurückbringen. Wir bezeichnen mit X_i den Winkel, um den sich die Kreisscheibe bei der i-ten Umdrehung $(i = 1, 2, \ldots)$ dreht. Ferner bezeichnen wir mit Y_n die Häufigkeit des Auftretens des Ereignisses A in den ersten n Umdrehungen. Man zeige, daß (Robbins [3])

$$P\left[\lim_{n\to\infty} Y_n = \frac{\delta_1}{\delta_1 + \delta_2}\right] = 1$$

gilt, wenn die Zufallsvariablen X_i $(i = 1, 2, \ldots)$ eine beliebige gemeinsame Verteilungsfunktion mit Ausnahme der Möglichkeit haben, daß die X_i eine gitterförmige Verteilung mit einem rationalen Vielfachen von π für die maximale Schrittweite $h > 0$ aufweisen.

Wir weisen darauf hin, daß auf Grund des starken Gesetzes der großen Zahlen von Borel

$$P\left[\lim_{n\to\infty} Y_n = P(A)\right] = 1$$

wäre, falls wir nach jeder Umdrehung die Kreisscheibe in ihre Ausgangsstellung zurückgeführt hätten (siehe auch Aufgabe 14.6.8).

7. MARKOFFSCHE KETTEN

7.1. Einleitende Bemerkungen

Bis jetzt haben wir uns hauptsächlich mit unabhängigen zufälligen Ereignissen und unabhängigen Zufallsvariablen beschäftigt. In den Anwendungen der Wahrscheinlichkeitsrechnung kann man oft annehmen, daß die betrachteten zufälligen Ereignisse oder Zufallsvariablen unabhängig sind. Es gibt aber zahlreiche Probleme in Physik, Technik und anderen Anwendungsgebieten der Wahrscheinlichkeitsrechnung, in denen die Voraussetzung über die Unabhängigkeit nicht einmal näherungsweise erfüllt ist. Die Betrachtung abhängiger zufälliger Ereignisse und abhängiger Zufallsvariabler gehört mit zu den wichtigsten Aufgaben der Wahrscheinlichkeitsrechnung. Wenn aber nicht mehr Unabhängigkeit vorausgesetzt werden kann, werden die Überlegungen der Wahrscheinlichkeitsrechnung sehr viel schwieriger. Das große Verdienst MARKOFFS [2] besteht nun darin, daß er bei der Untersuchung von abhängigen Ereignissen ein Versuchsschema hervorhob (man sagt jetzt, dieses Versuchsschema bilde eine Markoffsche Kette), das als einfachste Verallgemeinerung des unabhängigen Versuchsschemas betrachtet werden kann. Die Markoffschen Untersuchungen waren der Ausgangspunkt einer weitgehenden Entwicklung, die zur Entdeckung eines wichtigen neuen Gebiets der modernen Wahrscheinlichkeitsrechnung führte, nämlich zur *Theorie der stochastischen Markoffschen Prozesse*. Einige Elemente dieser Theorie sind Gegenstand der Betrachtungen dieses und des nächsten Kapitels.

Die Markoffschen Ketten, von denen in diesem Kapitel die Rede ist, sind ebenfalls, wie wir in 8.2 sehen werden, Markoffsche Prozesse; aus didaktischen und historischen Gründen hielten wir es aber für angebracht, die Markoffschen Ketten gesondert zu behandeln.

7.2. Homogene Markoffsche Ketten

Wir setzen voraus, daß alle in diesem und im nächsten Kapitel vorkommenden bedingten Wahrscheinlichkeiten definiert sind (siehe 2.7).

Wir betrachten eine Folge von Versuchen; als Ergebnis eines jeden Versuchs möge ein und nur ein Ereignis der endlich oder abzählbar vielen Ereignisse $E_1, E_2, E_3, \ldots$ eintreten können; dabei sollen die Ereignisse $E_1, E_2, \ldots$ einander paarweise ausschließen. Wir wollen diese Ereignisse *Zustände* oder *Phasen* nennen. Tritt das Ereignis E_j ein, so wollen wir sagen, das System befinde sich im Zustand oder in der Phase E_j. Das Symbol $E_j^{(k)}$ gebrauchen wir, um auszudrücken, daß der

Zustand E_j beim k-ten Versuch realisiert wurde, während das Symbol $E_j^{(0)}$ bedeutet, daß E_j die Anfangsphase des Systems war. Weiter bezeichnen wir mit $p_{ij}^{(k)}$ die bedingte Wahrscheinlichkeit dafür, daß sich das System nach dem k-ten Versuch im Zustand E_j befindet, wenn es sich beim $(k-1)$-ten im Zustand E_i befand, d. h.

$$p_{ij}^{(k)} = P\left(E_j^{(k)} \mid E_i^{(k-1)}\right).$$

Definition 7.2.1. Wir sagen, eine Folge von Versuchen bilde eine *Markoffsche Kette*, wenn für beliebige $i, j, k = 1, 2, 3, \ldots$ die Gleichung

$$p_{ij}^{(k)} = P\left(E_j^{(k)} \mid E_i^{(k-1)}\right) = P\left(E_j^{(k)} \mid E_i^{(k-1)}, E_{i_{k-2}}^{(k-2)}, \ldots, E_{i_1}^{(1)}, E_{i_0}^{(0)}\right) \tag{7.2.1}$$

für beliebige

$$E_{i_{k-2}}^{(k-2)}, \ldots, E_{i_1}^{(1)}, E_{i_0}^{(0)}$$

erfüllt ist.

Definition 7.2.2. Wir sagen, eine Versuchsfolge bilde eine *homogene Markoffsche Kette*, wenn für $i, j, k = 1, 2, 3, \ldots$ die Wahrscheinlichkeiten $p_{ij}^{(k)}$ nicht von k abhängen:

$$p_{ij}^{(k)} = p_{ij} \qquad (k = 1, 2, \ldots). \tag{7.2.2}$$

Die Wahrscheinlichkeit p_{ij} heißt *Übergangswahrscheinlichkeit* vom Zustand E_i in den Zustand E_j während eines Versuches.

Manchmal wollen wir eine zeitliche Terminologie gebrauchen, d. h., wir denken uns, daß während jeder Zeiteinheit ein Versuch ausgeführt wird, und sagen an Stelle von „beim k-ten Versuch geht das System vom Zustand E_i in den Zustand E_j über" auch „der Übergang findet im Augenblick $t = k$ statt". Außerdem nehmen wir an, daß zu Beginn, d. h. im Moment $t = 0$, sich das System mit der Wahrscheinlichkeit $P(E_i)$ im Zustand E_i befindet. Nach dieser Bezeichnungsweise ist p_{ij} die Wahrscheinlichkeit des Übergangs vom Zustand E_i in den Zustand E_j während einer Zeiteinheit.

Aus den Formeln (7.2.1), (7.2.2) und (1.5.7) erhalten wir für eine homogene Markoffsche Kette die Formel

$$\begin{aligned} P(E_{i_0} E_{i_1} \cdots E_{i_n}) &= P(E_{i_0})\, P(E_{i_1} \mid E_{i_0}) \cdots P(E_{i_n} \mid E_{i_{n-1}}) \\ &= P(E_{i_0})\, p_{i_0 i_1} \cdots p_{i_{n-1} i_n}, \end{aligned} \tag{7.2.3}$$

die die Wahrscheinlichkeit des Produkts $E_{i_0} E_{i_1} E_{i_2} \cdots E_{i_n}$ der Zustände in n aufeinanderfolgenden Versuchen angibt.

Der Leser vergleiche die letzte Formel mit der Formel (1.5.7) und beachte den wesentlichen Unterschied zwischen ihnen.

Aus (7.2.3) folgt, daß die Wahrscheinlichkeit eines beliebigen Produkts von Zuständen dann gegeben ist, wenn man alle Übergangswahrscheinlichkeiten p_{ij} und alle Anfangswahrscheinlichkeiten $P(E_i)$ für $i = 1, 2, 3, \ldots$ kennt.

7.3. Die Übergangsmatrix

A. Die Matrix, deren Elemente die Übergangswahrscheinlichkeiten p_{ij} sind, heißt die *Übergangsmatrix der Markoffschen Kette.* Wir wollen sie mit M_1 bezeichnen:

$$M_1 = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

Da alle Elemente p_{ij} Wahrscheinlichkeiten darstellen, sind sie nicht negativ.

Das System befinde sich im Zustand E_i. Das Ereignis, das darin besteht, daß sich das System nach einem Versuch entweder im Zustand E_i oder in einem der Zustände E_j, $j \neq i$, befindet, ist ein sicheres Ereignis. Da jedoch die Ereignisse E_j einander paarweise ausschließen, folgt für $i = 1, 2, 3, \ldots$ die Gleichung

$$P\Big(\big(\sum_j E_j\big) \,|\, E_i\Big) = \sum_j p_{ij} = 1. \qquad (7.3.1)$$

Die Summe aller Wahrscheinlichkeiten in jeder Zeile der Matrix M_1 ist also gleich Eins. Die Summe aller Wahrscheinlichkeiten einer Spalte braucht aber nicht gleich Eins zu sein.

Beispiel 7.3.1. Wir betrachten die Folge von Versuchen eines Bernoullischen Schemas. Hier haben wir zwei Zustände, E_1 und E_2, und für jeden Versuch ist

$$p_{11} = p_{21} = p,$$

$$p_{12} = p_{22} = q.$$

Die Übergangsmatrix hat also die Form

$$M_1 = \begin{pmatrix} p & q \\ p & q \end{pmatrix}.$$

Wie man leicht nachprüft, sind für jede Folge von unabhängigen Versuchen alle Zeilen der Übergangsmatrix einander gleich.

Beispiel 7.3.2. In diesem Beispiel betrachten wir die sogenannte Irrfahrt eines Moleküls zwischen zwei Absorptionsschirmen. Das ist ein Modell für gewisse Erscheinungen, welche in der Physik häufig vorkommen. Das Molekül kann sich in einem der Punkte $1, 2, 3, \ldots, s$ auf der x-Achse befinden; dabei bleibt es mit der Wahrscheinlichkeit Eins im Punkt 1, wenn es sich dort im Augenblick t befand. Dasselbe gilt für den Punkt s. Die Punkte 1 und s nennt man *Absorptionsschirme.* Befindet sich aber das Molekül im Moment t im Punkt i, wobei

$2 \leqq i \leqq s - 1$ ist, dann kann es während einer Zeiteinheit mit der Wahrscheinlichkeit p zum Punkt $i + 1$ oder mit der Wahrscheinlichkeit $q = 1 - p$ zum Punkt $i - 1$ übergehen. Hier haben wir es mit einer homogenen Markoffschen Kette mit s Zuständen zu tun. Dabei besteht der Zustand E_i darin, daß sich das Molekül im Punkt i befindet. Tatsächlich hängt die Übergangswahrscheinlichkeit vom Zustand i (den das Molekül im Augenblick t inne hat) in den Zustand j (den das Molekül im Augenblick $t + 1$ inne hat) nur vom Zustand i, in dem sich das Molekül im Augenblick t befindet, ab, nicht aber von den Zuständen, in denen sich das Molekül in der Zeit vor t befand. Die Übergangswahrscheinlichkeit hängt auch nicht unmittelbar von der Zeit t ab (sie hängt nur vom Zustand im Augenblick t ab). Wir haben hier die Übergangswahrscheinlichkeiten

$$p_{11} = p_{ss} = 1$$

und für $2 \leqq i \leqq s - 1$

$$p_{ij} = \begin{cases} p & \text{für } j = i + 1, \\ q = 1 - p & \text{für } j = i - 1, \\ 0 & \text{für die übrigen } j. \end{cases}$$

Die Übergangsmatrix M_1 hat die Form

$$M_1 = \begin{pmatrix} 1\,0\,0 \ldots\ldots 0 \\ q\,0\,p \ldots\ldots 0 \\ 0\,q\,0\,p \ldots\ldots 0 \\ \ldots\ldots\ldots\ldots \\ 0\,0\,0\,0 \ldots q\,0\,p \\ 0\,0\,0\,0 \ldots 0\,0\,1 \end{pmatrix}$$

Beispiel 7.3.3. In der auf den Mendelschen Gesetzen aufgebauten Genetik nimmt man an, daß die vererbbaren Merkmale von Genen abhängen. Die Gene treten immer zu Paaren auf. Im einfachsten Fall, auf welchen wir uns hier beschränken wollen, soll jedes Gen eine der zwei Formen A oder a besitzen. Wenn beide Gene eines betrachteten Organismus von der Form A sind, sagen wir, daß der Organismus die genotypische Struktur AA besitzt; falls aber ein Gen von der Form A, das andere jedoch von der Form a ist, sprechen wir von der genotypischen Struktur Aa und endlich, falls beide von der Form a sind, von der genotypischen Struktur aa. Weiter nimmt man an, daß die Keimzellen nur ein Gen enthalten. Also enthalten die Keimzellen eines Organismus von der genotypischen Struktur AA bzw. aa das Gen A bzw. a, während die Keimzelle eines Organismus mit der genotypischen Struktur Aa das Gen A oder a, beide mit der gleichen Wahrscheinlichkeit, enthalten kann. Ein Nachkomme erhält von jedem seiner Eltern ein Gen, und zwar nach dem Bernoullischen Schema. Das ist so zu verstehen: Man betrachtet die Gesamtheit der Gene aller zu derselben Generation gehörigen Individuen, zu der die Eltern des betrachteten Nachkommen gehören. Aus dieser Gesamtheit lost man nach dem Bernoullischen Schema zwei Gene aus. Ebenso erhalten wir die genotypische Struktur von N Nachkommen als Ergebnis von $2N$ derartigen Auslosungen.

Die Individuenanzahl sei nun in allen Generationen gleich N. Das kann man dadurch erreichen, daß man aus jeder Generation eine entsprechende Anzahl von Individuen herausgreift. In einer Generation befinden sich also $2N$ Gene. Haben in einer gewissen Generation $i\,(0 \leqq i \leqq 2N)$ unter den $2N$ Genen die Gestalt A, so sagen wir, die Gesamtheit befinde sich

im Zustand E_i. Aus dem angenommenen Fortpflanzungsschema ersieht man, daß man es mit einer homogenen Markoffschen Kette mit $2N + 1$ möglichen Zuständen $E_0, E_1, \ldots, E_{2N}$ zu tun hat. Die Übergangswahrscheinlichkeiten vom Zustand E_i, in dem sich die eine Generation befindet, in den Zustand E_j, in dem sich die nächste Generation befindet, sind durch

$$p_{ij} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}$$

gegeben.

Hier sehen wir, daß die Zustände E_0 und E_{2N} wie Absorptionsschirme wirken. Befindet sich nämlich die Gesamtheit einer Generation in einem dieser beiden Zustände, dann verläßt sie ihn nicht mehr. Haben z. B. alle Individuen die genotypische Struktur AA, dann kann kein Nachkomme Gene von der Gestalt a haben (vgl. MALÉCOT [1]).

Beispiel 7.3.4. In diesem Beispiel betrachten wir eine Irrfahrt ohne Absorptionsschirme. Die Menge der Zustände sei die Menge der natürlichen Zahlen, und die Übergangswahrscheinlichkeiten seien durch die folgenden Formeln gegeben:

$$p_{11} = q = 1 - p,$$

$$p_{ij} = \begin{cases} p & \text{für} \quad i = 1, 2, 3, \ldots, \; j = i + 1, \\ q & \text{für} \quad i = 2, 3, \ldots, \; j = i - 1, \\ 0 & \text{für die übrigen Paare } (i, j). \end{cases}$$

Die Zahl Eins nennt man *Reflexionsschirm*.

Die Übergangsmatrix hat die Form

$$M_1 = \begin{pmatrix} q\,p\,0\,0\,0\,0 \ldots \\ q\,0\,p\,0\,0\,0 \ldots \\ 0\,q\,0\,p\,0\,0 \ldots \\ \ldots\ldots\ldots\ldots \end{pmatrix}.$$

Beispiel 7.3.5. Wir kehren in diesem Beispiel zum Pólyaschen Schema zurück und wollen die Bezeichnungen aus 5.4 benutzen. Hier gibt es zwei Zustände, nämlich den Zustand E_1, der im Ziehen einer weißen Kugel besteht, und den Zustand E_2, der im Ziehen einer schwarzen Kugel besteht. Dabei sind die entsprechenden Anfangswahrscheinlichkeiten gleich

$$p_1 = \frac{b}{N}, \quad p_2 = 1 - p_1 = \frac{c}{N}.$$

Die Übergangswahrscheinlichkeit dafür, aus dem Zustand E_1 bei der zweiten Ziehung in den Zustand E_1 zu gelangen, ist gleich $\frac{b + s}{N + s}$. Dagegen ist die Wahrscheinlichkeit dafür, daß wir beim dritten Mal eine weiße Kugel erhalten, wenn beim zweiten Mal E_1 eintrat, gleich $\frac{b + 2s}{N + 2s}$, wenn bei der ersten Ziehung der Zustand E_1 eintrat, und gleich $\frac{b + s}{N + 2s}$, wenn bei der ersten Ziehung der Zustand E_2 eintrat. Die Folge der Versuche eines Pólyaschen Schemas bildet also keine Markoffsche Kette.

Aber wir können auch aus dem Pólyaschen Versuchsschema eine Markoffsche Kette bilden; dazu müssen wir nur die Zustände anders definieren. Wir wollen sagen, das System

befinde sich nach k Ziehungen im Zustand E_i $(i = 0, 1, 2, \ldots, k)$, wenn man bei diesen k Ziehungen i weiße Kugeln erhalten hat. Dann kann sich das System bei der $(k + 1)$-ten Ziehung entweder wieder im Zustand E_i befinden oder aber zum Zustand E_{i+1} übergehen. Das hängt nur davon ab, ob wir bei der $(k + 1)$-ten Ziehung eine schwarze oder eine weiße Kugel erhalten haben. Die Übergangswahrscheinlichkeiten sind dabei von den Ergebnissen der ersten $k - 1$ Versuche unabhängig, sie sind jedoch von der Versuchsnummer abhängig. Hier haben wir es also mit einer inhomogenen Markoffschen Kette zu tun, deren Übergangswahrscheinlichkeiten $p_{ij}^{(k+1)}$ durch

$$p_{ij}^{(k+1)} = \begin{cases} \dfrac{c + (k - i)\, s}{N + ks} & \text{für } j = i, \\ \dfrac{b + is}{N + ks} & \text{für } j = i + 1, \\ 0 & \text{für } j \neq i, i + 1 \end{cases}$$

gegeben werden.

B. Die Übergangswahrscheinlichkeit dafür, in einer homogenen Markoffschen Kette nach n Versuchen vom Zustand E_i in den Zustand E_j zu gelangen, wollen wir mit $p_{ij}(n)$ bezeichnen. Wir wollen zeigen, wie man $p_{ij}(n)$ durch p_{ij} ausdrücken kann, und beginnen mit der Berechnung von $p_{ij}(2)$. Das Ereignis A, das im Übergang des Systems vom Zustand E_i in den Zustand E_j besteht, ist gleich der Summe von Ereignissen A_k. Dabei bezeichnet A_k den Übergang vom Zustand E_i zu E_k und von E_k zu E_j. Also gilt für jedes Zahlenpaar (i, j)

$$p_{ij}(2) = \sum_k p_{ik} p_{kj}, \tag{7.3.2}$$

wobei die Summation über alle möglichen Zustände zu erstrecken ist.

Durch ähnliche Überlegungen erhält man für $n = 3, 4, \ldots$

$$p_{ij}(n) = \sum_k p_{ik}(m)\, p_{kj}(n - m); \tag{7.3.3}$$

dabei ist m eine natürliche Zahl, die der Ungleichung $1 \leqq m < n$ genügt. Die Gleichung (7.3.3) wird auch *Markoffsche Gleichung* genannt. Sie ist für die Theorie der homogenen Markoffschen Ketten von grundlegender Bedeutung.

Die Matrix, deren Elemente die Übergangswahrscheinlichkeiten $p_{ij}(n)$ sind, wollen wir *Übergangsmatrix in n Schritten* nennen und mit dem Symbol M_n bezeichnen. Die Abhängigkeit der Matrizen M_n und M_1 läßt sich leicht angeben. Zuerst wollen wir die Abhängigkeit der Matrizen M_1 und M_2 bestimmen. Aus der Formel (7.3.2) erkennt man, daß das Element der Matrix M_2, das zur i-ten Zeile und j-ten Spalte gehört, gleich der Summe der Produkte der Elemente der i-ten Zeile mit den Elementen der j-ten Spalte der Matrix M_1 ist. Nach der Multiplikationsregel für Matrizen erhalten wir

$$M_2 = M_1^2.$$

Wenn wir vollständige Induktion anwenden und die Formel (7.3.3) berücksichtigen, erhalten wir

$$M_n = M_1^n. \tag{7.3.4}$$

7.4. Ein Ergodensatz

A. Diesen Paragraphen beginnen wir mit einer von KOLMOGOROFF [9] stammenden Klassifikation der Zustände einer Markoffschen Kette (vgl. auch die Aufgaben 7.6.5 bis 7.6.8). Das soll uns die Deutung der Voraussetzungen des Ergodensatzes erleichtern.

Definition 7.4.1. Wir nennen einen Zustand E_i *vorübergehend*, wenn es einen Zustand E_j und eine natürliche Zahl k gibt derart, daß $p_{ij}(k) > 0$ und $p_{ji}(m) = 0$ für $m = 1, 2, 3, \ldots$ ist.

Definition 7.4.2. Ein Zustand E_i heißt *wesentlich*, wenn aus der Existenz einer natürlichen Zahl k_j mit der Eigenschaft $p_{ij}(k_j) > 0$, die für einen gewissen Zustand E_j vorhanden sein soll, die Existenz einer natürlichen Zahl m_i mit der Eigenschaft $p_{ij}(m_i) > 0$ folgt.

Definition 7.4.3. Ein wesentlicher Zustand E_i heißt *periodisch*, wenn es eine natürliche Zahl $d > 1$ gibt, so daß für alle durch d nicht teilbaren n die Beziehung $p_{ii}(n) = 0$ gilt.

Im Beispiel 7.3.2 sind die Zustände E_1 und E_s wesentlich und nicht periodisch, während alle übrigen Zustände vorübergehend sind. Wir können nämlich vom Zustand E_i $(i = 2, \ldots, s-1)$ in $i-1$ Schritten mit positiver Wahrscheinlichkeit zum Zustand E_1 übergehen, während die Wahrscheinlichkeit dafür, aus dem Zustand E_1 herauszukommen, gleich Null ist. Wir bemerken noch, daß es nicht möglich ist, vom Zustand E_1 zum Zustand E_s überzugehen, ebenso ist auch der Übergang von E_s zu E_1 unmöglich, obwohl beide Zustände wesentlich sind. In diesem Zusammenhang bringen wir noch eine weitere Definition:

Definition 7.4.4. Wir sagen, eine Menge W von wesentlichen Zuständen bilde eine *einzige Klasse*, wenn es zu je zwei Zuständen E_i und E_j aus W eine natürliche Zahl m_{ij} mit der Eigenschaft $p_{ij}(m_{ij}) > 0$ gibt.

Die Beispiele 7.4.1 bis 7.4.4 werden die hier eingeführten Begriffe erläutern.

B. Wir wollen nun zur Darstellung des Ergodensatzes übergehen. Dieser Satz wird die Frage beantworten, wie sich die Wahrscheinlichkeiten $p_{ij}(n)$ für $n \to \infty$ verhalten. Mit anderen Worten: Hat der Anfangszustand E_i bei einer genügend großen Anzahl von Übergängen einen Einfluß auf das Eintreten des Ereignisses E_j? Der Satz 7.4.1 gibt eine Bedingung dafür an, daß bei einer homogenen Markoffschen Kette mit endlich vielen Zuständen die Wahrscheinlichkeit $p_{ij}(n)$ für $n \to \infty$ gegen eine von i unabhängige Zahl p_j konvergiert $(j = 1, 2, \ldots, s)$. Der Satz erschöpft nicht den gesamten Inhalt dieses Problems. So betrifft er zum

Beispiel nicht Markoffsche Ketten mit abzählbar vielen Zuständen. Besser informieren über dieses Gebiet kann sich der Leser in den Büchern von FELLER [7], CHUNG [5] sowie von KEMENEY und SNELL [1], ferner in den Arbeiten von CHUNG [2], [3] (vgl. auch Aufgabe 7.6.9).

Satz 7.4.1. *Es sei* $M_1 = (p_{ij})$ *die Übergangsmatrix einer homogenen Markoffschen Kette mit endlich vielen Zuständen* $E_1, \ldots, E_s$. *Wenn es eine natürliche Zahl* r *gibt derart, daß die Elemente* $p_{ij}(r)$ *der Matrix* M_r *in* s_1 *Spalten* $(s_1 \geqq 1)$ *der Relation*

$$\min_{1 \leqq i \leqq s} p_{ij}(r) = \delta > 0 \tag{7.4.1}$$

genügen, dann gelten die Beziehungen

$$\lim_{n \to \infty} p_{ij}(n) = p_j \qquad (j = 1, 2, \ldots, s); \tag{7.4.2}$$

hierbei ist $p_j \geqq \delta$ *für diejenigen* j, *für welche* (7.4.1) *gilt. Außerdem ist*

$$\sum_{j=1}^{s} p_j = 1$$

und

$$|p_{ij}(n) - p_j| \leqq (1 - s_1\delta)^{\frac{n}{r} - 1} \tag{7.4.3}$$

In den Voraussetzungen des Satzes wird gefordert, daß alle Elemente wenigstens einer Spalte der Matrix M_r positiv sind. Der Satz 7.4.1 stellt eine gewisse Modifikation eines Satzes von MARKOFF [2] dar, in welchem man fordert, daß für ein gewisses natürliches r alle Elemente $p_{ij}(r)$ der Matrix M_r positiv sind. Die Behauptung des Satzes lautet dann $p_j > 0$ $(j = 1, 2, \ldots, s)$.

Den Satz von 7.4.1 bezeichnet man als *Ergodensatz*. Diese Bezeichnung wollen wir am Ende von 7.4 erklären.

C. Bevor wir zum Beweis des Satzes 7.4.1 kommen, wollen wir einige Beispiele behandeln, die den Sinn der Voraussetzung dieses Satzes erläutern sollen.

Beispiel 7.4.1. Wir kehren zum Beispiel 7.3.2 zurück und setzen der Einfachheit halber $s = 3$. Es ist also

$$M_1 = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix}.$$

Daraus berechnen wir

$$M_2 = M_1^2 = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix} = M_1.$$

Allgemein ist also

$$M_n = M_1.$$

Die Voraussetzungen des Ergodensatzes sind hier nicht erfüllt. Es gibt kein r derart, daß M_r wenigstens eine Spalte mit lauter positiven Elementen $p_{ij}(r)$ hätte. Offenbar ist gleichzeitig die Behauptung des Satzes 7.4.1 nicht erfüllt; denn es ist $p_{11}(n) = 1$, also $\lim_{n\to\infty} p_{11}(n) = 1$, während $\lim_{n\to\infty} p_{21}(n) = q$ und $\lim_{n\to\infty} p_{31}(n) = 0$ ist. Die Unregelmäßigkeit dieser Markoffschen Kette ist durch die Existenz zweier wesentlicher Zustände E_1 und E_3 bedingt, die nicht in einer Klasse liegen; denn der Übergang von dem einen zum anderen ist unmöglich.

Beispiel 7.4.2. Wir betrachten eine homogene Markoffsche Kette mit den vier Zuständen E_1, E_2, E_3, E_4 und der Übergangsmatrix

$$M_1 = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Dann ist

$$M_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

und allgemein für $k = 1, 2, 3, \ldots$

$$M_{2k+1} = M_1,$$

$$M_{2k} = M_2.$$

Auch hier ist weder die Voraussetzung noch die Behauptung des Satzes 7.4.1 erfüllt. Der Leser beachte die in dieser Kette auftretende Periodizität. Alle Zustände sind wesentlich, aber periodisch. So kann z. B. das System von Zustand E_1 zum Zustand E_1 nur nach einer geraden Anzahl von Schritten mit positiver Wahrscheinlichkeit zurückkehren. Diese Periodizität ruft eine Unregelmäßigkeit hervor, so daß die Behauptung des Satzes 7.4.1 nicht gilt.

Beispiel 7.4.3. Wir kehren zum Beispiel 7.3.4 zurück. Dabei nehmen wir an, daß drei

Zustände vorhanden sind und die Matrix M_1 die Form

$$M_1 = \begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix}$$

hat. Dann ist

$$M_2 = \begin{pmatrix} q^2 + pq & qp & p^2 \\ q^2 & 2qp & p^2 \\ q^2 & pq & qp + p^2 \end{pmatrix}$$

Die Voraussetzungen des Satzes 7.4.1 sind also erfüllt. Wir weisen darauf hin, daß alle drei Zustände wesentlich und nicht periodisch sind und daß sie nur eine einzige Klasse bilden. Später werden wir zeigen, wie man für dieses Beispiel die Grenzwahrscheinlichkeiten berechnen kann.

Beispiel 7.4.4. Wir wollen das Beispiel 7.4.1 so ändern, daß die Übergangsmatrix M_1 die Form

$$M_1 = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix}$$

hat. In dem so gewonnenen Beispiel kann man den Zustand E_1 als Absorptions- und den Zustand E_3 als Reflexionsschirm auffassen. Es ist

$$M_2 = \begin{pmatrix} 1 & 0 & 0 \\ q & pq & p^2 \\ q^2 & pq & qp + p^2 \end{pmatrix}$$

Dann sind die Voraussetzungen des Satzes 7.4.1 erfüllt. Leicht erkennt man, daß der Zustand E_1 wesentlich und nicht periodisch ist, während die beiden anderen Zustände vorübergehend sind. Später zeigen wir (vgl. Beispiel 7.4.6), daß die Grenzwahrscheinlichkeiten p_2 und p_3 gleich Null sind.

Die obigen Beispiele legen die Vermutung nahe (es kann nachgewiesen werden, daß diese Vermutung tatsächlich zutrifft), daß die Voraussetzung des Satzes 7.4.1 erfüllt ist, wenn in einer homogenen Markoffschen Kette mit endlich vielen Zuständen alle wesentlichen Zustände nichtperiodisch sind und nur eine Klasse bilden. Dabei ist nicht ausgeschlossen, daß einige Zustände vorübergehend sind. Sind jedoch alle Zustände wesentlich und nichtperiodisch und bilden sie dann nur eine einzige Klasse, so gibt es ein r, für das alle Elemente $p_{ij}(r)$ der Matrix M_r positiv sind und damit auch größer als ein gewisses $\delta > 0$; denn die Anzahl der Elemente ist endlich. Es gilt dann der Markoffsche Satz im ursprünglichen Wortlaut.

Kaucky [7] und Konêcný [1] haben für Anwendungen günstige notwendige und hinreichende Bedingungen für die Gültigkeit des Ergodensatzes für homogene Markoffsche Ketten angegeben. Diese Bedingungen werden durch Eigenwerte der Matrix M_1 ausgedrückt.

D. Wir gehen jetzt zum Beweis des Satzes 7.4.1 über.

Beweis. Dazu führen wir die folgenden Bezeichnungen ein $(v = 1, 2, 3, \ldots)$:

$$b_j(v) = \min_{1 \leqq i \leqq s} p_{ij}(v),$$

$$B_j(v) = \max_{1 \leqq i \leqq s} p_{ij}(v). \tag{7.4.4}$$

Unter Berücksichtigung der Formel (7.3.3) erhalten wir für $v = 1, 2, \ldots$

$$b_j(v+1) = \min_{1 \leqq i \leqq s} p_{ij}(v+1) = \min_{1 \leqq i \leqq s} \sum_{k=1}^{s} p_{ik} p_{kj}(v)$$

$$\geqq \min_{1 \leqq i \leqq s} \sum_{k=1}^{s} p_{ik} b_j(v) = b_j(v),$$

also

$$b_j(v+1) \geqq b_j(v). \tag{7.4.5}$$

Ebenso erhalten wir

$$B_j(v+1) \leqq B_j(v). \tag{7.4.6}$$

Aus (7.4.5) und (7.4.6) folgt

$$b_j(1) \leqq b_j(2) \leqq \cdots \leqq B_j(2) \leqq B_j(1). \tag{7.4.7}$$

Die Zahl r erfülle die Voraussetzung unseres Satzes. Es seien i und m fest gewählt, und die Bezeichnungen $\sum_k^+$ sowie $\sum_k^-$ sollen besagen, daß die Summation über alle diejenigen k zu erstrecken ist, für die $p_{ik}(r) \geqq p_{mk}(r)$ bzw. $p_{ik}(r) < p_{mk}(r)$ ist. Dann erhalten wir

$$\sum_k^+ [p_{ik}(r) - p_{mk}(r)] + \sum_k^- [p_{ik}(r) - p_{mk}(r)] = 0. \tag{7.4.8}$$

Es sei nun $n > r$. Wir betrachten die Differenz

$$\begin{aligned} B_j(n) - b_j(n) &= \max_{1 \leqq i \leqq s} p_{ij}(n) - \min_{1 \leqq m \leqq s} p_{mj}(n) \\ &= \max_{1 \leqq i \leqq s} \sum_{k=1}^{s} p_{ik}(r)\, p_{kj}(n-r) - \min_{1 \leqq m \leqq s} \sum_{k=1}^{s} p_{mk}(r)\, p_{kj}(n-r) \\ &= \max_{1 \leqq i, m \leqq s} \sum_{k=1}^{s} [p_{ik}(r) - p_{mk}(r)]\, p_{kj}(n-r) \\ &\leqq \max_{1 \leqq i, m \leqq s} \Big\{ \sum_k^+ [p_{ik}(r) - p_{mk}(r)]\, B_j(n-r) \\ &\qquad + \sum_k^- [p_{ik}(r) - p_{mk}(r)]\, b_j(n-r) \Big\}. \end{aligned}$$

Wenn wir die Formel (7.4.8) berücksichtigen, so erhalten wir

$$B_j(n) - b_j(n) \leqq \max_{1 \leqq i, m \leqq s} \sum_k^+ [p_{ik}(r) - p_{mk}(r)] [B_j(n-r) - b_j(n-r)]$$
$$= [B_j(n-r) - b_j(n-r)] \max_{1 \leqq i, m \leqq s} \sum_k^+ [p_{ik}(r) - p_{mk}(r)]. \tag{7.4.9}$$

Die Relation (7.4.1) bestehe für w Glieder der Summe $\sum_k^+$. Offensichtlich ist $w \leqq s_1$, wenn s_1 die Anzahl der Spalten bedeutet, für die (7.4.1) gilt. Es ist also

$$-\sum_k^+ p_{mk}(r) \leqq -w\delta.$$

Da ferner für $s_1 - w$ Glieder der Summe $\sum_k^-$ die Beziehung (7.4.1) erfüllt ist, erhalten wir

$$\sum_k^+ p_{ik}(r) \leqq 1 - (s_1 - w)\,\delta \leqq 1$$

und daraus

$$\sum_k^+ [p_{ik}(r) - p_{mk}(r)] \leqq 1 - (s_1 - w)\,\delta - w\delta = 1 - s_1\delta. \tag{7.4.10}$$

Aus den Formeln (7.4.9) und (7.4.10) folgt die Ungleichung

$$B_j(n) - b_j(n) \leqq (1 - s_1\delta)\,[B_j(n-r) - b_j(n-r)].$$

Ebenso ist (für $n > 2r$)

$$B_j(n) - b_j(n) \leqq (1 - s_1\delta)^2\,[B_j(n-2r) - b_j(n-2r)].$$

Wenn wir das $\left[\frac{n}{r}\right]$-mal[1]) wiederholen, erhalten wir

$$B_j(n) - b_j(n) \leqq (1 - s_1\delta)^{\left[\frac{n}{r}\right]} \left\{B_j\left(n - \left[\frac{n}{r}\right] r\right) - b_j\left(n - \left[\frac{n}{r}\right] r\right)\right\}. \tag{7.4.11}$$

Hieraus folgt, da $\delta > 0$ und $s_1 \geqq 1$ ist, die Ungleichung

$$0 \leqq 1 - s_1\delta < 1.$$

Auf Grund der Formel (7.4.7) konvergieren die Folgen $\{b_j(n)\}$ und $\{B_j(n)\}$,

[1]) Unter $[A]$ verstehen wir die größte ganze Zahl, die nicht größer als A ist.

während sich aus (7.4.11) ergibt, daß diese Folgen einen gemeinsamen Grenzwert haben. Also ist

$$\lim_{n\to\infty} \max_{1\leq i\leq s} p_{ij}(n) = \lim_{n\to\infty} \min_{1\leq i\leq s} p_{ij}(n) = p_j. \tag{7.4.12}$$

Die Formel (7.4.2) ist damit bewiesen. Ferner besteht offenbar für diejenigen j, für die die Beziehung (7.4.1) gilt, die Ungleichung $p_j \geqq \delta$. Die Gleichung $\sum_{j=1}^{s} p_j = 1$ ist ebenfalls leicht einzusehen.

Uns bleibt nur noch die Formel (7.4.3) zu beweisen. Wir erhalten sie aus (7.4.7) und (7.4.11):

$$|p_{ij}(n) - p_j| \leqq B_j(n) - b_j(n) \leqq (1 - s_1\delta)^{\frac{n}{r}-1}.$$

Damit haben wir den Satz 7.4.1 vollständig bewiesen.

E. Jetzt wollen wir zeigen, wie man die Grenzwahrscheinlichkeiten p_j berechnet, falls sie existieren. Aus der Formel (7.3.3) ergibt sich

$$p_{ij}(n) = \sum_{k=1}^{s} p_{ik}(n-1)\, p_{kj}.$$

Existieren die Grenzwahrscheinlichkeiten p_j, so erhalten wir, wenn wir auf beiden Seiten der letzten Gleichung für $n \to \infty$ zur Grenze übergehen, das Gleichungssystem

$$p_j = \sum_{k=1}^{s} p_k p_{kj} \qquad (j = 1, 2, \ldots, s). \tag{7.4.13}$$

Aus diesen Gleichungen und der Beziehung

$$\sum_{j=1}^{s} p_j = 1$$

kann man die gesuchten Wahrscheinlichkeiten p_j bestimmen.

Beispiel 7.4.5. Wir kehren zum Beispiel 7.4.3 zurück und berechnen die Grenzwahrscheinlichkeiten p_j. Die Formel (7.4.13) ergibt drei lineare Gleichungen

$$p_1 = p_1 q + p_2 q,$$

$$p_2 = p_1 p + p_3 q,$$

$$p_3 = p_2 p + p_3 p.$$

Daraus erhalten wir

$$p_2 = \frac{p}{q}\, p_1,$$

$$p_3 = \left(\frac{p}{q}\right)^2 p_1.$$

Aus der Beziehung $p_1 + p_2 + p_3 = 1$ erhalten wir

$$p_1 \left[1 + \left(\frac{p}{q}\right) + \left(\frac{p}{q}\right)^2\right] = 1.$$

Für $p = q = \frac{1}{2}$ ist $p_1 = p_2 = p_3 = \frac{1}{3}$. Dann sind im Limes alle Zustände gleich wahrscheinlich. Ist aber $p \neq q$, so erhalten wir

$$p_j = \frac{1 - \frac{p}{q}}{1 - \left(\frac{p}{q}\right)^3} \left(\frac{p}{q}\right)^{j-1} \qquad (j = 1, 2, 3).$$

Wenn $p > q$ ist, wachsen die Wahrscheinlichkeiten p_j mit wachsendem j, während für $q > p$ das entgegengesetzte Verhalten auftritt. Diese Ergebnisse stimmen mit der Anschauung überein. So ist z. B. für $\frac{p}{q} = 2$

$$p_1 = \frac{1}{7}, \quad p_2 = \frac{2}{7}, \quad p_3 = \frac{4}{7},$$

während wir für $\frac{p}{q} = \frac{1}{2}$

$$p_1 = \frac{4}{7}, \quad p_2 = \frac{2}{7}, \quad p_3 = \frac{1}{7}$$

erhalten. Wir sehen, daß alle drei Grenzwahrscheinlichkeiten positiv sind. Das hängt damit zusammen, daß sie wesentlich sind (bei einer Kette mit abzählbar vielen Zuständen braucht das nicht der Fall zu sein; vgl. auch Aufgabe 7.6.9).

Beispiel 7.4.6. Wir wollen die Grenzwahrscheinlichkeit p_j des Beispiels 7.4.4 berechnen. Aus (7.4.13) erhalten wir

$$p_1 = p_1 + p_2 q,$$

$$p_2 = p_3 q,$$

$$p_3 = (p_2 + p_3)\, p.$$

Daraus und aus der Beziehung $p_1 + p_2 + p_3 = 1$ folgt $p_1 = 1$, $p_2 = p_3 = 0$.

Wie wir schon vorhin erwähnt haben, liegt das daran, daß die Zustände E_2 und E_3 vorübergehend sind.

Den Begriff der Ergodizität und die Bedingungen für die Gültigkeit des Ergodensatzes für inhomogene Markoffsche Ketten findet man in den Arbeiten von Kolmogoroff [3], Sarymsakow [1] und Hajnal [1], [2].

F. Wir betrachten jetzt die Beziehungen, die zwischen den Grenzwahrscheinlichkeiten und den totalen Wahrscheinlichkeiten in einer homogenen Markoffschen

Kette bestehen. Wir berechnen die totale Wahrscheinlichkeit dafür, daß sich das System nach n Übergängen im Zustand E_j befindet. Diese Wahrscheinlichkeit wollen wir mit $c_j(n)$ bezeichnen. Es ist

$$c_j(n) = \sum_k P(E_k)\, p_{kj}(n) = \sum_k c_k(n-1)\, p_{kj}, \qquad (7.4.14)$$

wobei $P(E_k)$ die Wahrscheinlichkeit dafür ist, daß sich das System zu Beginn im Zustand E_k befindet.

Definition 7.4.5. Wir nennen eine homogene Markoffsche Kette *stationär*, wenn die Gleichungen

$$P(E_j) = c_j(1) \qquad (j = 1, 2, \ldots)$$

erfüllt sind. Die $c_j(n)$ heißen dann *stationäre totale Wahrscheinlichkeiten.*

Aus den letzten Gleichungen folgen für $j = 1, 2, \ldots$ und $n = 1, 2, 3, \ldots$ die Gleichungen

$$c_j(1) = c_j(2) = \cdots = c_j(n) = c_j.$$

Aus der Formel (7.4.14) erhalten wir also

$$c_j = \sum_k c_k p_{kj} \qquad (j = 1, 2, \ldots). \qquad (7.4.15)$$

Wir nehmen nun an, daß die Anzahl der Zustände endlich und gleich s ist und daß die Voraussetzungen des Satzes 7.4.1 erfüllt sind; die Grenzwahrscheinlichkeiten p_j existieren also. Wie man leicht erkennt [man vergleiche nur die Formeln (7.4.13) und (7.4.15) miteinander], bestehen dann die Gleichungen $c_j = p_j$ $(j = 1, 2, \ldots, s)$. Wenn die Anfangswahrscheinlichkeiten $P(E_j)$ gleich den Grenzwahrscheinlichkeiten p_j sind, dann sind die totalen Wahrscheinlichkeiten $c_j(n)$ konstant, und zwar gleich p_j. Die Kette ist also stationär, und es herrscht Gleichgewicht in dem Sinne, daß die Verteilung der totalen Wahrscheinlichkeiten konstant ist. Daraus erklärt sich die Bezeichnung *Ergodensatz.*

Wir bemerken noch, daß für eine beliebige Markoffsche Kette mit endlich vielen Zuständen der folgende Satz über die Grenzwerte der totalen Wahrscheinlichkeiten gilt:

Satz 7.4.2. *Dafür, daß in einer homogenen Markoffschen Kette mit endlich vielen Zuständen die Grenzwerte der totalen Wahrscheinlichkeiten*

$$\lim_{n\to\infty} c_j(n) = c_j \qquad (j = 1, 2, \ldots, s) \qquad (7.4.16)$$

existieren und von der Anfangsverteilung unabhängig sind, ist die Existenz der ergodischen Wahrscheinlichkeiten p_j notwendig und hinreichend; dabei ist $c_j = p_j$ $(j = 1, 2, \ldots, s)$.

Beweis. Es sei die Beziehung (7.4.16) erfüllt, und c_j sei von der Anfangsverteilung unabhängig. Wir können also als Anfangsverteilung $P(E_i) = 1$ und $P(E_j) = 0$ $(j \neq i)$ wählen. Die Formel (7.4.14) ergibt dann

$$c_j(n) = p_{ij}(n),$$

und wenn wir noch (7.4.16) berücksichtigen, erhalten wir

$$p_j = \lim_{n \to \infty} p_{ij}(n) = c_j \qquad (i, j = 1, 2, \ldots, s).$$

Jetzt setzen wir voraus, daß die ergodischen Wahrscheinlichkeiten p_j existieren. Für eine beliebige Anfangsverteilung erhalten wir dann aus (7.4.14)

$$\lim_{n \to \infty} c_j(n) = \lim_{n \to \infty} \sum_{k=1}^{s} P(E_k)\, p_{kj}(n) = p_j \sum_{k=1}^{s} P(E_k) = p_j,$$

was zu beweisen war.

7.5. Zufallsvariable, die eine homogene Markoffsche Kette bilden

A. Die Überlegungen der vorhergehenden Paragraphen lassen sich auf Zufallsvariable übertragen. Es seien X_k $(k = 0, 1, 2, \ldots)$ Zufallsvariable, von denen jede die Werte x_i $(i = 1, 2, 3, \ldots)$ annehmen kann. Die Werte x_i entsprechen den Zuständen E_i, von denen bisher die Rede war. Wir bringen jetzt Definitionen, die den Definitionen aus 7.2 analog sind.

Definition 7.5.1. Wir sagen, daß eine Folge $\{X_k\}$ $(k = 0, 1, 2, \ldots)$ von Zufallsvariablen mit den möglichen Zuständen x_i $(i = 1, 2, 3, \ldots)$ eine *Markoffsche Kette* bildet, wenn für $i, j, k = 1, 2, 3, \ldots$ die Gleichungen

$$p_{ij}^{(k)} = P(X_k = x_j \mid X_{k-1} = x_i) \tag{7.5.1}$$

$$= P(X_k = x_j \mid X_{k-1} = x_i,\; X_{k-2} = x_{i_{k-2}}, \ldots, X_1 = x_{i_1},\; X_0 = x_{i_0})$$

für beliebige $x_{i_{k-2}}, \ldots, x_{i_1}, x_{i_0}$ gelten.

Definition 7.5.2. Wir sagen, daß eine Folge $\{X_k\}$ $(k = 0, 1, 2, \ldots)$ von Zufallsvariablen eine *homogene* Markoffsche Kette bildet, wenn für $i, j, k = 1, 2, 3, \ldots$ die bedingten Wahrscheinlichkeiten $p_{ij}^{(k)}$ von k unabhängig sind:

$$p_{ij}^{(k)} = p_{ij}. \tag{7.5.2}$$

In die Redeweise der Zufallsvariablen übersetzt, ist die Übergangswahrscheinlichkeit $p_{ij}(n)$ die Wahrscheinlichkeit dafür, daß $X_n = x_j$ ist, wenn $X_0 = x_i$ war, also

$$p_{ij}(n) = P(X_n = x_j \mid X_0 = x_i).$$

Die Formel (7.3.3) hat hier die Gestalt

$$p_{ij}(n) = \sum_r P(X_m = x_r | X_0 = x_i)\, P(X_n = x_j | X_m = x_r), \tag{7.5.3}$$

wobei $1 \leq m < n$ ist.

Die durch (7.4.14) bestimmten totalen Wahrscheinlichkeiten $c_j(n)$ haben hier die Form

$$c_j(n) = P(X_n = x_j) = \sum_r P(X_0 = x_r)\, P(X_n = x_j | X_0 = x_r). \tag{7.5.4}$$

Die Definition 7.4.5 nimmt die folgende Gestalt an:

Definition 7.5.3. Eine Folge von Zufallsvariablen X_k $(k = 0, 1, \ldots)$, die eine homogene Markoffsche Kette bilden, heißt *stationär*, wenn für $j = 1, 2, 3, \ldots$ die Gleichungen

$$c_j(0) = P(X_0 = x_j) = P(X_1 = x_j) = c_j(1) = c_j$$

gelten.

Aus der letzten Gleichung folgt, daß für $k = 0, 1, 2, \ldots$ und $j = 1, 2, 3, \ldots$ die Gleichungen

$$P(X_k = x_j) = c_j$$

erfüllt sind. In einer stationären Folge von Zufallsvariablen haben also alle Zufallsvariablen die gleiche Verteilung.

Ebenso leicht lassen sich die Klassifikation der Zustände und der oben bewiesene Satz in der Sprache der Zufallsvariablen formulieren. Das überlassen wir jedoch dem Leser.

B. Es muß gezeigt werden, daß die Theorie der Grenzverteilungen für Zufallsvariable, die eine homogene Markoffsche Kette bilden, wesentlich weniger entwickelt ist als die gleiche Theorie für unabhängige Zufallsvariable.

Die Bedingungen für die Gültigkeit des zentralen Grenzwertsatzes für Markoffsche Ketten mit drei Zuständen wurden von Markoff [3] gefunden; für Ketten mit einer beliebigen endlichen Anzahl von Zuständen wurden diese Bedingungen von Romanowsky [1], Fréchet [2], Onicescu und Mihoc [1], [2] angegeben. Doeblin [2] zeigte, daß für gewisse Klassen von Markoffschen Ketten mit einer abzählbaren Menge von Zuständen die Frage der Grenzwertsätze sich auf die entsprechende Frage für unabhängige Zufallsvariable zurückführen läßt. Für Ketten mit beliebig vielen Zuständen wurden gewisse Ergebnisse von Doeblin [3], Doob [5], Dynkin [2] und Chung [2], [3] erzielt. Ein verhältnismäßig weitreichendes Ergebnis, das im wesentlichen eine Verallgemeinerung des Satzes von Lindeberg-Lévy für eine umfangreiche Klasse von Zufallsvariablen darstellt, die eine homogene Markoffsche Kette bilden, wurde kürzlich von Nagajew [1]

gewonnen. Ein lokaler Grenzwertsatz für Markoffsche Ketten mit endlich vielen Zuständen wurde von KOLMOGOROFF [13] angegeben. SIRASHDINOW [1] erzielte gewisse Ergebnisse über das Konvergenzverhalten bezüglich der Grenzverteilung im lokalen und integralen Grenzwertsatz für Markoffsche Ketten mit endlich vielen Zuständen. In der genannten Arbeit entwickelt NAGAJEW den zentralen lokalen Grenzwertsatz für Markoffsche Ketten mit abzählbar vielen Zuständen und schätzt die Konvergenz gegen die Normalverteilung ab.

Gewisse Sätze bezüglich des Gesetzes der großen Zahlen für Zufallsvariable, die eine Markoffsche Kette bilden, findet man in dem Buch von DOOB [5] und in der Arbeit von CHUNG [3]. Allgemeine Ergebnisse auf diesem Gebiet fand kürzlich BREIMAN [2].

Wir wollen uns hier damit begnügen, das Gesetz der großen Zahlen und den zentralen Grenzwertsatz für Zufallsvariable, die eine homogene Markoffsche Kette mit endlich vielen Zuständen bilden, ohne Beweis anzugeben.

Satz 7.5.1. *Es sei* $\{X_k\}$ $(k = 0, 1, 2, \ldots)$ *eine stationäre Folge von Zufallsvariablen, die eine homogene Markoffsche Kette mit endlich vielen Zuständen bilden. Sind alle wesentlichen Zustände nichtperiodisch und bilden sie eine einzige Klasse, dann besteht die Beziehung*

$$P\left(\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} X_k = E(X_0)\right) = 1. \tag{7.5.5}$$

Sind also die Voraussetzungen dieses Satzes erfüllt, so gilt für die Folge $\{X_k\}$ das starke Gesetz der großen Zahlen. Der Leser vergleiche diesen Satz mit dem Satz 6.12.3 von KOLMOGOROFF.

Beispiel 7.5.1. Wir betrachten eine stationäre Folge $\{X_k\}$ von Zufallsvariablen mit einer Zweipunktverteilung, die eine homogene Markoffsche Kette bilden. Dem Zustand E_1 ordnen wir die Zahl x_1 und dem Zustand E_2 die Zahl x_2 zu.

Die Übergangsmatrix hat die Form

$$M_1 = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

wobei $0 < p_{12} < 1$, $0 < p_{21} < 1$ ist. Die Voraussetzung des Satzes 7.4.1 ist hier schon für $r = 1$ erfüllt. Beide Zustände sind wesentlich und nichtperiodisch und bilden nur eine einzige Klasse. Aus der Formel (7.4.13) und der Beziehung $p_1 + p_2 = 1$ finden wir für die ergodischen Wahrscheinlichkeiten

$$p_1 = \frac{p_{21}}{p_{12} + p_{21}},$$

$$p_2 = \frac{p_{12}}{p_{12} + p_{21}}; \tag{7.5.6}$$

dabei ist $0 < p_1 < 1$, $0 < p_2 < 1$. Aus der Voraussetzung, daß die Folge $\{X_k\}$ stationär ist,

folgt $P(X_k = x_1) = p_1$ und $P(X_k = x_2) = p_2$ $(k = 0, 1, 2, \ldots)$. Nach dem Satz 7.5.1 gilt mit der Wahrscheinlichkeit Eins die Beziehung

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} X_k = \frac{p_{21}}{p_{12}+p_{21}} x_1 + \frac{p_{12}}{p_{12}+p_{21}} x_2. \tag{7.5.7}$$

Insbesondere sei $x_1 = 1$ und $x_2 = 0$. Das Ereignis $\left(\sum_{k=0}^{n} X_k = m\right)$ besteht dann darin, daß sich das System bei $n + 1$ Versuchen m-mal im Zustand E_1 befindet, und die Relation (7.5.7) besagt: Mit der Wahrscheinlichkeit Eins konvergiert das arithmetische Mittel der Anzahl Male, in denen sich das System im Zustand E_1 befindet, gegen $\dfrac{p_{21}}{p_{12}+p_{11}}$. Wenn wir also das Erscheinen des Wertes $x_1 = 1$ als Erfolg bezeichnen, so sehen wir, daß für die Anzahl der Erfolge in einer homogenen stationären Markoffschen Kette das starke Gesetz der großen Zahlen gilt. Für dieses Schema hatte schon Markoff [1] das schwache Gesetz der großen Zahlen erhalten.

Wir formulieren jetzt den oben erwähnten Grenzwertsatz. Dazu setzen wir

$$Y_n = \sum_{k=0}^{n} [X_k - E(X_k)].$$

Satz 7.5.2. *Die Zufallsvariable X_k $(k = 0, 1, 2, \ldots)$ mögen eine homogene Markoffsche Kette mit endlich vielen Zuständen bilden. Sind alle Zustände wesentlich, bilden sie nur eine einzige Klasse und erfüllt die Dispersion $D^2(Y_n)$ die Bedingung*

$$\lim_{n\to\infty} \frac{D^2(Y_n)}{n+1} = \sigma^2 > 0, \tag{7.5.8}$$

so ist unabhängig von der Anfangsverteilung der Zufallsvariablen X_0 die Beziehung

$$\lim_{n\to\infty} P\left(\frac{Y_n}{\sigma\sqrt{n+1}} < y\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{y^2}{2}}\, dy \tag{7.5.9}$$

erfüllt.

Beispiel 7.5.2. Wir kehren zum Beispiel 7.5.1 zurück. Dort betrachteten wir die stationäre Folge $\{X_k\}$. Es sei $x_1 = 1$ und $x_2 = 0$. Wir setzen $Z_k = X_k - E(X_k)$. Wegen der Stationarität ist $E(X_k) = p_1$ $(k = 0, 1, 2, \ldots)$. Wir wollen $D^2(Y_n)$ berechnen und nachprüfen, ob die Beziehung (7.5.8) erfüllt ist. Wenn wir wieder berücksichtigen, daß die Folge $\{X_k\}$ stationär ist, so erhalten wir

$$\begin{aligned} D^2(Y_n) &= \sum_{k=0}^{n} D^2(Z_k) + 2\sum_{k=0}^{n-1} \sum_{m=k+1}^{n} E(Z_k Z_m) \\ &= (n+1)\, p_1 p_2 + 2\sum_{k=0}^{n-1} \sum_{m=k+1}^{n} E(Z_k Z_m). \end{aligned} \tag{7.5.10}$$

Um $E(Z_k Z_m)$ zu berechnen, beachten wir, daß die Zufallsvariable $Z_k Z_m$ die folgenden Werte mit den entsprechenden Wahrscheinlichkeiten annehmen kann:

$$P\left(Z_k Z_m = (1-p_1)^2\right) = P(X_k = 1)\, P(X_m = 1 \mid X_k = 1) = p_1 p_{11}(m-k);$$

$$\begin{aligned} P\left(Z_k Z_m = (1-p_1)(-p_1)\right) &= P(X_k = 1)\, P(X_m = 0 \mid X_k = 1) \\ &\quad + P(X_k = 0)\, P(X_m = 1 \mid X_k = 0) \\ &= p_1 p_{12}(m-k) + p_2 p_{21}(m-k); \end{aligned}$$

$$P(Z_k Z_m = p_1^2) = P(X_k = 0)\, P(X_m = 0 \mid X_k = 0) = p_2 p_{22}(m-k).$$

Durch einfache Umformungen erhalten wir damit

$$\begin{aligned} E(Z_k Z_m) &= p_1 p_2^2 p_{11}(m-k) - p_1^2 p_2 p_{12}(m-k) - p_1 p_2^2 p_{21}(m-k) + p_1^2 p_2 p_{22}(m-k) \\ &= p_1 p_2 [p_{11}(m-k) - p_{21}(m-k)] = p_1 p_2 (p_{11} - p_{21})^{m-k}. \end{aligned} \tag{7.5.11}$$

Wenn wir die Bezeichnung $\delta = p_{11} - p_{21}$ einführen, erhalten wir aus der letzten Beziehung und aus (7.5.10)

$$\begin{aligned} D^2(Y_n) &= p_1 p_2 \left[n + 1 + 2[n\delta + (n-1)\delta^2 + \cdots + \delta^n]\right] \\ &= p_1 p_2 \left[n + 1 + 2\left(\sum_{j=1}^{n} \delta^j + \sum_{j=1}^{n-1} \delta^j + \cdots + \delta\right)\right] \\ &= p_1 p_2 \left[n + 1 + \frac{2\delta n}{1-\delta} - \frac{2\delta^2(1-\delta^n)}{(1-\delta)^2}\right]. \end{aligned}$$

Daraus folgt

$$\frac{D^2(Y_n)}{n+1} = p_1 p_2 \left(1 + 2\frac{\delta}{1-\delta}\frac{n}{n+1}\right) - \frac{p_1 p_2}{n+1}\frac{2\delta^2(1-\delta^n)}{(1-\delta)^2}.$$

Schließlich erhalten wir, wenn wir berücksichtigen, daß $0 < \delta < 1$ ist,

$$\lim_{n\to\infty} \frac{D^2(Y_n)}{n+1} = p_1 p_2 \frac{1+\delta}{1-\delta} = \sigma^2 > 0.$$

Die Bedingung (7.5.8) ist also erfüllt, und es gilt die Beziehung (7.5.9).

Die exakten Verteilungen der Anzahl erfolgreicher Versuche in homogenen Markoffschen Ketten mit zwei Zuständen, die wir in diesem Beispiel betrachtet hatten, und die Momente dieser Zufallsvariablen, wurden von Gabriel [1] untersucht. In dieser Arbeit findet der Leser auch eine Anwendung einer solchen Kette auf Probleme der Meteorologie.

Wir bemerken noch, daß aus (7.5.11) folgt, daß der Korrelationskoeffizient ϱ_{km} der Zufallsvariablen X_k und X_m durch

$$\varrho_{km} = (p_{11} - p_{21})^{m-k}$$

gegeben wird.

Das hier betrachtete Beispiel für die Gültigkeit des zentralen Grenzwertsatzes

stellt einen Spezialfall eines Satzes von MARKOFF [3] dar. Der Satz von MARKOFF gilt für Zufallsvariable, die eine inhomogene Markoffsche Kette bilden. Der Markoffsche Satz wurde von BERNSTEIN [3] verschärft. Ein wichtiges Ergebnis in diesem Problemkreis erhielt DOBRUSCHIN [3].

Wir verweisen den Leser auf neuere Arbeiten von HUNT [1], DOOB [10] sowie KEMENY und SNELL [2], [3], in denen neue Methoden zur Untersuchung von Markoffschen Ketten entwickelt werden; diese Methoden beruhen auf einer Verbindung zwischen den Markoffschen Ketten und der Potentialtheorie.

7.6. Aufgaben und Ergänzungen

1. Die Übergangsmatrix einer homogenen Markoffschen Kette mit vier Zuständen habe die Form

$$M_1 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{5} & 0 & \frac{1}{3} & \frac{7}{15} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

a) Man klassifiziere die Zustände.

b) Man stelle fest, ob der Ergodensatz gilt.

c) Gilt der Ergodensatz, so bestimme man die ergodischen Wahrscheinlichkeiten.

2. Man beantworte die oben angegebenen Fragen für die Matrix

$$M_1 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

3. (Das Modell von EHRENFEST). In der Physik, bei Problemen der Diffusion sowie der Periodizität und Reversibilität (eingehende Informationen hierüber findet man in dem interessanten Buch von KAC [3]) findet das folgende Modell einer homogenen Markoffschen Kette Anwendung: $2R$, mit den Nummern $1, 2, \ldots, 2R-1, 2R$ versehene Kugeln

sind in zwei Urnen, A und B, untergebracht. In jeder Zeiteinheit wird zufällig die Nummer einer Kugel gezogen, und zwar entsprechend dem Polynomialschema mit der konstanten Wahrscheinlichkeit $(2R)^{-1}$, eine der Zahlen $1, 2, \ldots, 2R$ zu ziehen. Sobald die Nummer einer Kugel gezogen wurde, geht die Kugel aus der Urne, in der sie sich während des Ziehungsvorganges befand, in die andere Urne über. Wir wollen sagen, daß das System der beiden Urnen sich in dem Zustand E_j $(j = 0, 1, \ldots, 2R)$ befindet, wenn die Urne A j Kugeln enthält. Damit haben wir eine homogene Markoffsche Kette mit $2R + 1$ Zuständen und der Übergangsmatrix

$$M_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \ldots 0 & 0 \\ (2R)^{-1} & 0 & 1-(2R)^{-1} & 0 & 0 & \ldots 0 & 0 \\ 0 & 2(2R)^{-1} & 0 & 1-2(2R)^{-1} & 0 & \ldots 0 & 0 \\ 0 & 0 & 3(2R)^{-1} & 0 & 1-3(2R)^{-1} & \ldots 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & \ldots 1 & 0 \end{pmatrix}.$$

a) Man zeige, daß die ergodischen Wahrscheinlichkeiten $p_j(a)$ existieren; sie sind demnach den stationären Wahrscheinlichkeiten c_j gleich, und es ist dabei

$$c_j = \binom{2R}{j} \frac{1}{2^{2R}} \qquad (j = 0, 1, \ldots, 2R).$$

b) Man beweise, daß unabhängig vom Anfangszustand des Systems sich im Grenzfall stets die Binomialverteilung ergibt.

c) Unter Verwendung der Aufgabe 5.14.2 stelle man fest, welcher Zustand im Grenzfall der wahrscheinlichste ist.

4. a) Man zeige, daß für jede homogene Markoffsche Kette mit endlich vielen Zuständen der Grenzwert

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}(k) = q_{ij}$$

existiert.

b) Man berechne q_{ij} in den Beispielen 7.4.1 bis 7.4.4.

5. (Fortsetzung der Kolmogoroffschen [9] Klassifizierung der Zustände). Es möge $K_{ij}(n)$ die Übergangswahrscheinlichkeit vom Zustand E_i in den Zustand E_j zum ersten Mal auf dem n-ten Schritt bedeuten, und es sei

$$L_{ij} = \sum_{n=1}^{\infty} K_{ij}(n), \qquad R_{ij} = \sum_{n=1}^{\infty} n K_{ij}(n).$$

Den Ausdruck R_{jj} nennt man die *mittlere Rückkehrzeit des Zustandes* E_j.
Man nennt den Zustand E_j *rekurrent*, wenn $L_{jj} = 1$ ist, und *transient*, wenn $L_{jj} < 1$ ist. Ein rekurrenter Zustand, für den $R_{jj} = \infty$ ist, heißt *Nullzustand*. Ein rekurrenter Zustand, der weder periodisch noch ein Nullzustand ist, wird als *ergodischer Zustand* bezeichnet.

Man beweise die folgenden Behauptungen:

a) $K_{ij}(n) = p_{ij}(n) - K_{ij}(1)\, p_{jj}(n-1) - \cdots - K_{ij}(n-1)\, p_{jj}$.

b) In einer homogenen Markoffschen Kette mit endlich vielen Zuständen ist der Zustand E_j dann und nur dann rekurrent (transient), wenn er wesentlich (unwesentlich) ist.

6. Wir betrachten eine homogene Markoffsche Kette mit abzählbar vielen Zuständen und der Übergangsmatrix

$$M_1 = \begin{pmatrix} p_1 & 1-p_1 & 0 & 0 & 0 \dots \\ p_2 & 0 & 1-p_2 & 0 & 0 \dots \\ p_3 & 0 & 0 & 1-p_3 & 0 \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Man beweise: Ist $\sum_{j=1}^{\infty} p_{ij} < \infty$, so sind alle Zustände transient, ist jedoch $\sum_{j=1}^{\infty} p_{ij} = \infty$, so sind alle Zustände rekurrent.

Man vergleiche dieses Ergebnis mit der Aufgabe 5b und folgere hieraus, daß Zustände existieren können, die sowohl transient als auch wesentlich sind.

7. Wir bezeichnen mit Ω_{ij} die Wahrscheinlichkeit dafür, daß ein System unendlich oft in den Zustand E_j zurückkehrt, wenn es den Zustand E_i als Ausgangszustand hatte. Man beweise die folgenden Behauptungen:

a) Ist $L_{jj} = 1$, so ist $\Omega_{jj} = 1$.

b) In einer Menge wesentlicher Zustände, die eine Klasse bilden, sind entweder alle $\Omega_{ij} < 1$ oder alle $\Omega_{ij} = 1$.

8. Man beweise, daß in einer Menge von wesentlichen Zuständen, die eine Klasse bilden, entweder alle $L_{ij} < 1$ oder alle $L_{ij} = 1$ sind.

9. (Verallgemeinerung des Ergodensatzes.) Es sei $M_1 = (p_{ij})$ die Übergangsmatrix einer homogenen Markoffschen Kette mit abzählbar vielen Zuständen $E_1, E_2, \dots$ Man beweise, daß dann die folgenden Behauptungen gelten:

a) Sind alle Zustände rekurrent, keine Nullzustände und nichtperiodisch und bilden sie eine Klasse, so gilt für $i, j = 1, 2, \dots$

$$\lim_{n\to\infty} p_{ij}(n) = p_j = \frac{1}{R_{jj}}$$

mit $p_1 + p_2 + \cdots = 1$, wobei $p_j > 0$ und $p_j = c_j$ ist; dabei ist c_j die stationäre Wahrscheinlichkeit.

b) Ist der Zustand E_j ein transienter oder rekurrenter Nullzustand, so ist $\lim_{n\to\infty} p_{ij}(n) = 0$.

c) Ist der Zustand E_j ein rekurrenter periodischer Nullzustand mit der Periode $d > 1$, so gilt

$$\lim_{n\to\infty} p_{ij}(n) = \frac{d}{R_{jj}}$$

(Kolmogoroff [9], Feller [7]).

10. Wir betrachten eine homogene Markoffsche Kette mit abzählbar vielen Zuständen und der Übergangsmatrix

$$M_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & \dots \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \dots \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & \dots \\ \frac{4}{5} & 0 & 0 & 0 & \frac{1}{5} & 0 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

Man zeige, daß dann $\lim\limits_{n\to\infty} p_{ij}(n) = p_j = \frac{e^{-1}}{j!}$ $(i, j = 1, 2, \dots)$ gilt. Man vergleiche hierzu Aufgabe 6.

11. Wir betrachten eine homogene Markoffsche Kette mit abzählbar vielen Zuständen und der Übergangsmatrix

$$M_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & \dots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & \dots \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 & \dots \\ \frac{1}{4} & 0 & 0 & 0 & \frac{3}{4} & 0 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

Man beweise, daß sämtliche Zustände rekurrente Nullzustände sind und somit

$$\lim_{n\to\infty} p_{ij}(n) = 0 \qquad (i, j = 1, 2, \dots)$$

gilt.

8. STOCHASTISCHE PROZESSE

8.1. Der Begriff des stochastischen Prozesses

In diesem Kapitel wollen wir den bisherigen Bereich der Wahrscheinlichkeitsrechnung erweitern. Diese Erweiterung erfolgt in zwei Richtungen. Wir werden nicht nur wie bisher Folgen von Zufallsvariablen betrachten, d. h. nur endliche oder abzählbare Mengen von Zufallsvariablen, sondern auch überabzählbare Mengen von Zufallsvariablen. Um die zweite Ausdehnung unseres Bereiches klar darstellen zu können, müssen wir noch einige einleitende Bemerkungen machen.

Wir haben in den vorangegangenen Kapiteln Zufallsvariable betrachtet, die jedem Elementarereignis ein n-Tupel von Zahlen zuordneten. Mit anderen Worten: Die Realisierung dieser Zufallsvariablen ist ein n-Tupel von Zahlen. Auch eine Folge von Zufallsvariablen $X_1, X_2, X_3, \ldots$ kann man als eine einzige Zufallsvariable auffassen. Ihre Realisierung ist dann die Zahlenfolge $(x_1, x_2, x_3, \ldots)$

In der Praxis treffen wir aber noch viel kompliziertere, vom Zufall abhängige Vorgänge an, deren Realisierungen reelle Funktionen sein können. So ist z. B. die Anzahl der Telefongespräche in Warschau während des Zeitabschnitts $(0, t)$ für festes t eine Zufallsvariable, während die Anzahl der Gespräche, als Funktion von t betrachtet (wobei t irgendein Intervall durchläuft), eine Zufallsfunktion ist.

Der Verbrauch an elektrischer Energie in Warschau zu einem gegebenen Zeitpunkt ist eine Zufallsvariable, dagegen ist der Energieverbrauch während einer längeren Zeit, als Funktion der Zeit betrachtet, eine Zufallsfunktion.

Ebenso ist der Wasserstand der Weichsel unter der Poniatowski-Brücke in Warschau, gemessen täglich in einem bestimmten Augenblick, eine Zufallsvariable-während der Wasserstand der Weichsel, der täglich in einem bestimmten Augen, blick im Abstand t von der Poniatowski-Brücke gemessen wird, als Funktion von t betrachtet, eine Zufallsfunktion ist.

Während die eine Erweiterung der Wahrscheinlichkeitsrechnung darin bestehen wird, daß wir jetzt auch nichtabzählbare Mengen von Zufallsvariablen betrachten, besteht die andere in der Untersuchung von Zufallsfunktionen. Die so erweiterte Wahrscheinlichkeitsrechnung nennt man *Theorie der stochastischen Prozesse.*

Definition 8.1.1. Wir nennen eine Menge von Zufallsvariablen X_t, die von einem Parameter t abhängen, der in einer gewissen reellen Zahlenmenge I variiert, einen *stochastischen Prozeß.*

Einen stochastischen Prozeß wollen wir mit $\{X_t, t \in I\}$ bezeichnen. Den Parameter t wollen wir die *Zeit* nennen.

Im Sinne dieser Definition ist eine k-dimensionale ($k \geqq 1$) Zufallsvariable ein stochastischer Prozeß, I ist hier die Menge $(1, 2, \ldots, k)$. Auch jede abzählbare Folge $\{X_k\}$ $(k = 1, 2, \ldots)$ von Zufallsvariablen ist ein stochastischer Prozeß; hier ist I die Menge aller natürlichen Zahlen. In diesen beiden Fällen sprechen wir von *Prozessen mit diskreter Zeit*. Eine wirkliche Erweiterung der Wahrscheinlichkeitsrechnung liegt aber erst dann vor, wenn I eine nichtabzählbare Menge ist. Dabei wollen wir uns auf den Fall beschränken, daß I ein endliches oder unendliches Intervall ist. Wir sprechen dann von einem *stochastischen Prozeß mit stetiger Zeit*.

Wir bemerken nur, daß man einen stochastischen Prozeß auch noch anders definieren kann. Man kann nämlich einen stochastischen Prozeß als eine Menge von Zufallsfunktionen auffassen, die den Elementarereignissen zugeordnet wurden. Hier begnügen wir uns damit, darauf hinzuweisen, daß X_t für festes t eine Zufallsvariable ist, während für festgewählte Elementarereignisse die Größe X_t eine Zufallsfunktion des Arguments t ist; diese Funktion wollen wir eine *Realisierung* des stochastischen Prozesses nennen.

Bei vielen Problemen aus der Theorie der stochastischen Prozesse bedeutet der Parameter t tatsächlich die Zeit, in der die untersuchte Erscheinung abläuft. Der Parameter t kann aber auch irgendeine andere physikalische Größe bezeichnen, z. B. den Raum, in dem der Prozeß verläuft, oder er braucht überhaupt keinen physikalischen Sinn zu haben.

8.2. Markoffsche Prozesse und Prozesse mit unabhängigen Zuwächsen

A. Wie wir schon im vorigen Kapitel bemerkten, entsteht die Notwendigkeit, sich mit abhängigen Zufallsvariablen zu beschäftigen, durch die Bedürfnisse der Physik und der Technik. Jetzt, da wir zur Untersuchung von nichtabzählbaren Mengen von Zufallsvariablen übergehen, müssen wir uns von vornherein mit abhängigen Zufallsvariablen beschäftigen. Setzen wir nämlich voraus, daß die Zufallsvariablen unabhängig sind, so können wir, wenn es überabzählbar viele sind, schwerlich erwarten, eine Gesetzmäßigkeit in diesem Chaos von zufälligen Ereignissen zu finden; andererseits ist die Untersuchung der Gesetzmäßigkeiten, die bei zufälligen Ereignissen auftreten, die Hauptaufgabe der Wahrscheinlichkeitsrechnung.

Wir wollen nun Definitionen von verschiedenen stochastischen Prozessen bringen und mit der Definition eines Markoffschen Prozesses beginnen. Diese Prozesse spielen eine große Rolle in der Theorie und den Anwendungen.

Definition 8.2.1. Einen stochastischen Prozeß $\{X_t, t \in I\}$ bezeichnen wir als *Markoffschen Prozeß*, wenn für $n = 1, 2, 3, \ldots$ und beliebige Parameterwerte

$t_m \in I$ $(m = 0, 1, \ldots, n;$ $t_0 < t_1 < \cdots < t_n)$ sowie für beliebige reelle Zahlen y, x die Gleichung

$$P(X_{t_n} < y \mid X_{t_{n-1}} = x, X_{t_{n-2}} = x_{n-2}, \ldots, X_{t_1} = x_1, X_{t_0} = x_0)$$
$$= P(X_{t_n} < y \mid X_{t_{n-1}} = x) \tag{8.2.1}$$

für alle $x_{n-2}, \ldots, x_0$ gilt.

Gemäß dieser Definition hängt die bedingte Verteilung der Zufallsvariablen X_{t_n} unter der Bedingung, daß die Zufallsvariable $X_{t_{n-1}}$ den Wert x annimmt, von keiner Information über die Werte ab, die durch die Zufallsvariablen dieses Prozesses in den t_{n-1} vorausgehenden Zeitpunkten angenommen werden. Für zwei beliebige Parameterwerte $t_1 \in I$ und $t_2 \in I$ mit $t_1 < t_2$ definieren wir die bedingte Verteilungsfunktion $F(t_1, x, t_2, y)$ durch die Formel

$$F(t_1, x, t_2, y) = P(X_{t_2} < y \mid X_{t_1} = x). \tag{8.2.2}$$

Diese Verteilungsfunktion ist von dem Wert von X_t zum Zeitpunkt $t < t_1$ unabhängig.

Mit den Eigenschaften der durch diese Formel bestimmten Verteilungsfunktionen wollen wir uns in 8.8 näher beschäftigen.

Definition 8.2.2. Ein Markoffscher Prozeß $\{X_t, t \in I\}$ ist *homogen*, wenn für beliebige $t_1 \in I$ und $t_2 \in I$ $(t_1 < t_2)$ die Verteilungsfunktion (8.2.2) nur von der Differenz $t = t_2 - t_1$ abhängt. Wir schreiben in diesem Fall

$$F(t_1, x, t_2, y) = F(t, x, y). \tag{8.2.3}$$

Die Prozesse, von denen in den beiden folgenden Definitionen die Rede ist, haben große praktische und theoretische Bedeutung.

Definition 8.2.3. Wir sagen, daß $\{X_t, t \in I\}$ ein *Prozeß mit unabhängigen Zuwächsen* ist, wenn für $n = 2, 3, \ldots$ und beliebige $t_m \in I$ $(m = 0, 1, \ldots, n;$ $t_0 < t_1 < \cdots < t_n)$ die Zufallsvariablen $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ unabhängig sind.

Erfüllt ein Prozeß $\{X_t, t \in I\}$ mit unabhängigen Zuwächsen die Bedingung $P(X_{t_0} = b) = 1$, wobei t_0 der Anfangszeitpunkt und b eine Konstante ist, dann ist er ein Markoffscher Prozeß. Es gilt dann (mit der Wahrscheinlichkeit Eins) die Gleichung

$$X_{t_n} = b + \sum_{m=1}^{n} (X_{t_m} - X_{t_{m-1}});$$

X_{t_n} ist also eine Summe von $n + 1$ unabhängigen Zufallsvariablen.

Andererseits braucht aber ein Markoffscher Prozeß kein Prozeß mit unabhängigen Zuwächsen zu sein. In der Tat, es sei $P(X_{t_0} = 0) = 1$, und

$$\{X_t, t_0 \leqq t < a\}$$

sei ein Markoffscher Prozeß. Für Werte t_1 und t_2, die den Ungleichungen

$$t_0 < t_1 < t_2 < a$$

genügen, erhalten wir dann

$$P(X_{t_1} = x_1, X_{t_2} = x_1 + x_2) = P(X_{t_1} = x_1, X_{t_2} - X_{t_1} = x_2)$$
$$= P(X_{t_1} = x_1)\, P(X_{t_2} - X_{t_1} = x_2 \,|\, X_{t_2} = x_1).$$

Nehmen wir an, der betrachtete Prozeß habe unabhängige Zuwächse. Dann wäre

$$P(X_{t_1} = x_1, X_{t_2} - X_{t_1} = x_2) = P(X_{t_1} = x_1)\, P(X_{t_2} - X_{t_1} = x_2).$$

Die rechten Seiten der letzten und der vorletzten Gleichung brauchen aber nicht gleich zu sein.

Definition 8.2.4. Der Prozeß $\{X_t, t \in I\}$ erfährt *homogene* Zuwächse, wenn für beliebige $t_1, t_2 \in I$ $(t_1 < t_2)$ die Wahrscheinlichkeitsverteilung der Zufallsvariablen $X_{t_2} - X_{t_1}$ nur von $t = t_2 - t_1$ abhängt.

Ein Prozeß mit unabhängigen Zuwächsen, der ein homogener Markoffscher Prozeß ist, ist, wie man leicht nachweisen kann, ein Prozeß mit homogenen Zuwächsen.

B. In 8.3 bis 8.7 wollen wir Markoffsche Prozesse behandeln, deren Zufallsvariable diskret sind, also mit der Wahrscheinlichkeit Eins endlich oder abzählbar viele Werte annehmen. Wir nehmen der Einfachheit halber an (diese Annahme ist in den Anwendungen häufig erfüllt), daß die Sprungstellen der Verteilungsfunktionen in ganzen Zahlen i liegen. Wir wollen diese Zahlen *Zustände* oder *Phasen* nennen, während wir den Prozeß selbst als *diskret* bezeichnen wollen. Für diesen Fall ist es bequemer, den Prozeß durch die bedingte Wahrscheinlichkeitsfunktion zu charakterisieren, anstatt wie bisher durch die bedingte Verteilungsfunktion. Die Definitionen 8.2.1 und 8.2.2 erfahren dann dementsprechende Änderungen.

Definition 8.2.1'. Ein diskreter Prozeß $\{X_t, t \in I\}$ heißt ein *Markoffscher Prozeß*, wenn für $n = 1, 2, 3, \ldots$ und für beliebige Parameterwerte $t_m \in I$ $(m = 0, 1, \ldots, n;\ t_0 < t_1 < \cdots < t_n)$ sowie für beliebige ganze Zahlen i, j die Gleichungen

$$P(X_{t_n} = j \,|\, X_{t_{n-1}} = i, X_{t_{n-2}} = i_{n-2}, \ldots, X_{t_0} = i_0) = P(X_{t_n} = j \,|\, X_{t_{n-1}} = i) \tag{8.2.4}$$

für alle $i_0, \ldots, i_{n-2}$ gelten.

Für zwei beliebige Parameterwerte $t_1 \in I$ und $t_2 \in I$ mit $t_1 < t_2$ definieren wir die Funktion der Übergangswahrscheinlichkeit durch die Formel

$$p_{ij}(t_1, t_2) = P(X_{t_2} = j \mid X_{t_1} = i). \tag{8.2.5}$$

Die Funktion $p_{ij}(t_1, t_2)$ ist vom Wert des Prozesses X_t zum Zeitpunkt $t < t_1$ unabhängig.

Der Leser vergleiche diese Definition 8.2.1′ mit der Definition 7.2.1.

Die Funktionen $p_{ij}(t_1, t_2)$ genügen selbstverständlich den Beziehungen

$$p_{ij}(t_1, t_2) \geqq 0. \tag{8.2.6}$$

Außerdem gelten für beliebige $t_1, t_2 \in I$ $(t_1 < t_2)$ und alle i die Gleichungen

$$\sum_j p_{ij}(t_1, t_2) = 1. \tag{8.2.7}$$

Die Gleichung

$$p_{ij}(t_1, t_2) = \sum_k p_{ik}(t_1, s)\, p_{kj}(s, t_2), \tag{8.2.8}$$

wobei $t_1 < s < t_2$ ist, gilt für beliebige t_1, t_2 und s. Sie läßt sich vollkommen analog wie die Gleichung (7.3.3) herleiten. Die Gleichung (8.2.8) ist eine Verallgemeinerung der Markoffschen Gleichung und wird nach CHAPMAN und KOLMOGOROFF benannt.

Aus den Formeln (8.2.4), (8.2.5) und (1.5.7) erhalten wir für $n = 1, 2, \ldots$ und beliebige $t_m \in I$ $(m = 0, 1, \ldots, n)$, wobei $t_0 < t_1 < \cdots < t_n$ ist, die folgende Formel für die Wahrscheinlichkeitsfunktion der Zufallsvariablen $(X_{t_0}, X_{t_1}, \ldots, X_{t_n})$:

$$\begin{aligned} &P(X_{t_0} = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) \\ &= P(X_{t_0} = i_0)\, p_{i_0 i_1}(t_0, t_1) \cdots p_{i_{n-1} i_n}(t_{n-1}, t_n). \end{aligned} \tag{8.2.9}$$

Die Wahrscheinlichkeitsfunktion der Zufallsvariablen X_{t_0} und die Übergangswahrscheinlichkeitsfunktionen bestimmen die Wahrscheinlichkeitsverteilung der $(n + 1)$-dimensionalen Zufallsvariablen $(X_{t_0}, X_{t_1}, \ldots, X_{t_n})$ für $n = 1, 2, 3, \ldots$ und für beliebige Parameterwerte t.

Definition 8.2.2′. Ein diskreter Markoffscher Prozeß $\{X_t, t \in I\}$ heißt *homogen*, wenn für beliebige i, j sowie für beliebige $t_1, t_2 \in I$ $(t_1 < t_2)$ die Übergangswahrscheinlichkeitsfunktionen nur von der Differenz $t = t_2 - t_1$ abhängen. Wir schreiben in diesem Fall

$$p_{ij}(t_1, t_2) = p_{ij}(t). \tag{8.2.10}$$

Die Beziehungen (8.2.6) bis (8.2.8) haben dann die Form

$$p_{ij}(t) \geqq 0, \tag{8.2.11}$$

$$\sum_j p_{ij}(t) = 1, \tag{8.2.12}$$

$$p_{ij}(t + s) = \sum_k p_{ik}(t)\, p_{kj}(s). \tag{8.2.13}$$

C. Die Sätze 7.4.1 und 7.4.2 kann man auf homogene Markoffsche Prozesse mit endlich vielen Zuständen und stetiger Zeit übertragen. Die Beweise sind im Grunde genommen dieselben wie für Markoffsche Ketten. Man hat nur statt des Parameters n den Parameter t zu setzen. Deshalb bringen wir diese Sätze ohne Beweis.

Satz 8.2.1. *Es sei* $\{X_t, 0 \leqq t < \infty\}$ *ein homogener Markoffscher Prozeß mit endlich vielen Zuständen* $i = 0, 1, \ldots, s$. *Gibt es ein* t^* $(0 \leqq t^* < \infty)$ *derart, daß*

$$p_{ij}(t^*) > 0 \qquad (i, j = 0, 1, 2, \ldots, s)$$

ist, so existieren die Grenzwerte

$$\lim_{t\to\infty} p_{ij}(t) = p_j > 0 \qquad (i, j = 0, 1, 2, \ldots, s). \tag{8.2.14}$$

Wir führen folgende Bezeichnung ein:

$$c_j(t) = P(X_t = j) \qquad (j = 0, 1, \ldots);$$

$c_j(t)$ ist also die totale Wahrscheinlichkeit dafür, daß der Prozeß im Moment t im Zustand j ist. Für einen homogenen Markoffschen Prozeß besteht die Beziehung

$$c_j(t) = \sum_i c_i(0)\, p_{ij}(t). \tag{8.2.15}$$

Satz 8.2.2. *Es sei* $\{X_t, 0 \leqq t < \infty\}$ *ein homogener Markoffscher Prozeß mit endlich vielen Zuständen* $i = 0, 1, \ldots, s$. *Die Grenzwerte*

$$\lim_{t\to\infty} c_j(t) = c_j \qquad (j = 0, 1, \ldots, s) \tag{8.2.16}$$

existieren und sind unabhängig von der Anfangsverteilung im Augenblick $t = 0$ *genau dann, wenn die ergodischen Wahrscheinlichkeiten* p_j *existieren, die durch die Formeln* (8.2.14) *bestimmt sind. Dann ist* $p_j = c_j$ $(j = 0, 1, \ldots, s)$.

Fuchs [1] hat den Satz 8.2.2 auf homogene Markoffsche Prozesse mit beliebig vielen Zuständen verallgemeinert.

Mit der weiteren Untersuchung der Eigenschaften von allgemeinen diskreten Markoffschen Prozessen wollen wir uns in 8.7 beschäftigen. In den nächsten Paragraphen behandeln wir zunächst einige wichtige Spezialfälle solcher Prozesse.

8.3. Der Poissonsche Prozeß

In 5.5 behandelten wir die Poissonsche Verteilung. Dort zeigten wir, daß man die Poissonsche Verteilung als Grenzverteilung einer Folge von Binomialverteilungen erhalten kann, und in den Beispielen 5.5.1 bis 5.5.2 überzeugten wir uns von der guten Übereinstimmung einiger beobachteter Verteilungen mit der Poissonschen Verteilung. Jetzt sind wir in der Lage, diese Erscheinung eingehender zu untersuchen. Diesmal gehen wir aber anders an die Frage nach der Entstehung der Poissonschen Verteilung heran. Wir werden nämlich gewisse Bedingungen formulieren (sie stellen die mathematische Idealisierung beobachtbarer Eigenschaften der untersuchten Erscheinungen dar) und zeigen, daß unter diesen Bedingungen die Poissonsche Verteilung auftritt. Dieser neue Weg zur Lösung der Frage, wann eine Poissonsche Verteilung auftritt, kann auch bei vielen anderen Wahrscheinlichkeitsverteilungen eingeschlagen werden.

Kehren wir zum Beispiel 5.5.2 zurück. Dort sprachen wir über die Emission von α-Teilchen durch radioaktive Substanzen. Wir wollen weiter unten die zur Poissonschen Verteilung führenden Bedingungen formulieren und dabei jede der Bedingungen an Hand dieses Beispiels interpretieren. Man kann nämlich auf Grund vieler Beobachtungen mit guter Annäherung annehmen, daß die Bedingungen für dieses Beispiel erfüllt sind.

Es sei X_t die Anzahl der beobachteten zufälligen Ereignisse im Intervall $[0, t]$, wobei $0 \leqq t < \infty$ ist. Wir haben hier einen stochastischen Prozeß $\{X_t, 0 \leqq t < \infty\}$ vor uns; dabei kann die Zufallsvariable X_t für jedes t die ganzzahligen nichtnegativen Werte $i = 0, 1, 2, \ldots$ annehmen. Weiter kann für beliebige t_1 und t_2 $(t_1 < t_2)$ der Zuwachs $X_{t_2} - X_{t_1}$ gleich $0, 1, 2, 3, \ldots$ sein.

Bezeichnen wir das eben behandelte Ereignis als „Signal", so können wir einen solchen Prozeß „Signalprozeß" nennen. In konkreten Anwendungen kann es sich wirklich um einen Signalprozeß handeln, wie dies auch im Beispiel 8.3.1 der Fall ist, wo die Arbeit einer Telefonzentrale betrachtet wurde. Im hier behandelten Beispiel 5.5.2 wird ein „Signal" durch die Emission eines α-Teilchens dargestellt, und X_t bezeichnet die Anzahl der emittierten α-Teilchen während des Zeitabschnitts $[0, t)$.

Wir setzen voraus, daß die folgenden Bedingungen erfüllt sind:

I. *Der Prozeß* $\{X_t, 0 \leqq t < \infty\}$ *hat unabhängige Zuwächse.*

Die Bedingung I besagt also, daß die Anzahlen der Signale in sich nicht überschneidenden Zeitintervallen

$$[t_0, t_1), [t_1, t_2), \ldots, [t_{n-1}, t_n)$$

unabhängige Zufallsvariable sind.

In unserem Beispiel sind die Anzahlen der α-Teilchen, die während sich nicht überschneidender Zeitintervalle emittiert wurden, unabhängig.

II. *Der Prozeß* $\{X_t, 0 \leqq t < \infty\}$ *hat homogene Zuwächse.*

Bedingung II besagt also, daß die Wahrscheinlichkeit für das Auftreten einer bestimmten Anzahl von Signalen in Zeitintervallen gleicher Länge konstant ist.

In dem betrachteten Beispiel ist die Wahrscheinlichkeit für die Emission einer bestimmten Anzahl von α-Teilchen in Intervallen gleicher Länge konstant.

Wir führen für $\Delta t > 0$ folgende Bezeichnung ein:

$$W_i(\Delta t) = P(X_{\Delta t} - X_0 = i) \qquad (i = 0, 1, 2, \ldots). \tag{8.3.1}$$

$W_i(\Delta t)$ ist also die Wahrscheinlichkeit dafür, daß im Intervall der Länge Δt das betrachtete Signal i-mal vorkommt. Wegen Bedingung II hängt diese Wahrscheinlichkeit nur von Δt ab.

III. *Es bestehen die Beziehungen*

$$\lim_{\Delta t \to 0} \frac{W_1(\Delta t)}{\Delta t} = \lambda \qquad (\lambda > 0), \tag{8.3.2}$$

$$\lim_{\Delta t \to 0} \frac{1 - W_0(\Delta t) - W_1(\Delta t)}{\Delta t} = 0. \tag{8.3.3}$$

Die Beziehung (8.3.2) besagt, daß die Wahrscheinlichkeit für das Auftreten eines einzigen Signals in einem kleinen Intervall der Länge Δt gleich $\lambda \Delta t + o(\Delta t)$ ist, während die Beziehung (8.3.3) aussagt: Die Wahrscheinlichkeit, daß wenigstens zwei Signale auf einmal in einem Zeitintervall der Länge Δt auftreten, ist von der Ordnung $o(\Delta t)$. Die Bedingung III bedeutet also inhaltlich: Signale treten plötzlich und blitzschnell in kleinen Zeitabschnitten (asymptotisch gleich Null) auf, aber nur einzeln, und nicht etwa zu Paaren, Tripeln usw. Bei der betrachteten Emission von α-Teilchen kann man in Annäherung annehmen, daß die Bedingung III erfüllt ist.

Wir bringen nun die Definition eines homogenen Poissonschen Prozesses.

Definition 8.3.1. Ein diskreter stochastischer Prozeß $\{X_t, 0 \leqq t < \infty\}$ mit unabhängigen Zuwächsen und den Zuständen $i = 0, 1, 2, \ldots$ heißt ein *homogener Poissonscher Prozeß*, wenn für einen beliebigen Punkt t $(0 \leqq t < \infty)$ die Beziehung

$$P(X_t = i) = \frac{(\lambda t)^i}{i!} \exp(-\lambda t) \tag{8.3.4}$$

gilt, wobei $\lambda > 0$ ist.

Satz 8.3.1. *Ein stochastischer Prozeß* $\{X_t, 0 \leqq t < \infty\}$, *wobei* X_t *die Anzahl der Signale im Zeitabschnitt* $[0, t)$ *bedeutet, der die Bedingungen I bis III und die Gleichung* $P(X_0 = 0) = 1$ *erfüllt, ist ein homogener Poissonscher Prozeß.*

Beweis. Wenn wir die Bezeichnung (8.3.1) und die Bedingungen I und II berücksichtigen, erhalten wir für $\Delta t > 0$

$$W_0(t + \Delta t) = W_0(t)\, W_0(\Delta t), \tag{8.3.5}$$

$$W_i(t + \Delta t) = \sum_{k=0}^{i} W_{i-k}(t)\, W_k(\Delta t) \qquad (i = 1, 2, \ldots). \tag{8.3.6}$$

In der Tat, wenn im Abschnitt $[0, t + \Delta t)$ kein einziges Signal auftritt, dann kann auch weder im Abschnitt $[0, t)$ noch in $[t, t + \Delta t)$ ein Signal aufgetreten sein, und umgekehrt. Da die Bedingung I erfüllt sein soll, erhalten wir (8.3.5).

Tritt das Signal im Abschnitt $[0, t + \Delta t)$ genau i-mal auf, so kann das auf $i + 1$ verschiedene, sich paarweise ausschließende Arten geschehen: Das Signal kann nämlich $(i - k)$-mal im Intervall $[0, t)$ und k-mal im Intervall $[t, t + \Delta t)$ $(k = 0, 1, \ldots, i)$ auftreten. Daraus und aus Bedingung I erhalten wir (8.3.6).

Wir wollen jetzt (8.3.2) bzw. (8.3.3) in der Form

$$W_1(\Delta t) = \lambda \Delta t + o(\Delta t) \tag{8.3.7}$$

bzw.

$$1 - W_0(\Delta t) - W_1(\Delta t) = \sum_{k=2}^{\infty} W_k(\Delta t) = o(\Delta t) \tag{8.3.8}$$

schreiben. Aus diesen Formeln folgt die Gleichung

$$W_0(\Delta t) = 1 - \lambda \Delta t + o(\Delta t). \tag{8.3.9}$$

Mit (8.3.9) erhalten wir aus (8.3.5)

$$\frac{W_0(t + \Delta t) - W_0(t)}{\Delta t} = -\lambda W_0(t) - W_0(t)\, \frac{o(\Delta t)}{\Delta t}.$$

Ebenso erhalten wir aus den Formeln (8.3.7) bis (8.3.9) und (8.3.6):

$$\frac{W_i(t + \Delta t) - W_i(t)}{\Delta t} = -\lambda W_i(t) + \lambda W_{i-1}(t) + \frac{o(\Delta t)}{\Delta t} \sum_{k=0}^{i} W_k(t).$$

Gehen wir in diesen Gleichungen zur Grenze für $\Delta t \to 0$ über, so finden wir das lineare Differentialgleichungssystem

$$W_0'(t) = -\lambda W_0(t), \tag{8.3.10}$$

$$W_i'(t) = -\lambda W_i(t) + \lambda W_{i-1}(t) \qquad (i = 1, 2, \ldots). \tag{8.3.11}$$

Berücksichtigen wir noch, daß $W_0(0) = 1$ ist, so ergibt sich aus der Gleichung (8.3.10) sofort

$$W_0(t) = \exp(-\lambda t). \tag{8.3.12}$$

Um das Gleichungssystem (8.3.11) zu lösen, setzen wir

$$W_i(t) = U_i(t) \exp(-\lambda t) \qquad (i = 0, 1, 2, \ldots). \tag{8.3.13}$$

Aus (8.3.12) erhalten wir $U_0(t) = 1$. Weiter ist

$$W_i'(t) = [U_i'(t) - \lambda U_i(t)] \exp(-\lambda t) \qquad (i = 1, 2, \ldots).$$

Daraus und aus (8.3.11) ergibt sich

$$\begin{aligned}[U_i'(t) - \lambda U_i(t)] \exp(-\lambda t) &= -\lambda[W_i(t) - W_{i-1}(t)] \\ &= -\lambda[U_i(t) - U_{i-1}(t)] \exp(-\lambda t) \qquad (i = 1, 2, \ldots).\end{aligned}$$

Die Funktionen $U_i(t)$ genügen also den Differentialgleichungen

$$U_i'(t) = \lambda U_{i-1}(t) \qquad (i = 1, 2, \ldots).$$

Wir integrieren und erhalten

$$U_i(t) - U_i(0) = \lambda \int_0^t U_{i-1}(t)\, dt \qquad (i = 1, 2, \ldots).$$

Aus der Definition der Funktion $U_i(t)$ ergibt sich $U_i(0) = W_i(0)$ für alle i. Also folgt aus $W_0(0) = 1$, $W_i(0) = 0$ für $i \geqq 1$, daß $U_i(0) = 0$ für alle $i \geqq 1$ ist, und wir erhalten schließlich

$$U_i(t) = \lambda \int_0^t U_{i-1}(t)\, dt \qquad (i = 1, 2, \ldots).$$

Es ist also

$$U_i(t) = \frac{(\lambda t)^i}{i!}.$$

Daraus und aus (8.3.13) folgen die Gleichungen

$$W_i(t) = \frac{(\lambda t)^i}{i!} \exp(-\lambda t) \qquad (i = 1, 2, \ldots). \tag{8.3.14}$$

Aus den Gleichungen (8.3.12) und (8.3.14) folgt, daß für einen beliebigen Punkt t $(0 \leqq t < \infty)$ die folgenden Gleichungen gelten:

$$P(X_t = i) = \frac{(\lambda t)^i}{i!} \exp(-\lambda t) \qquad (i = 0, 1, \ldots).$$

Der Satz 8.3.1 ist damit bewiesen.

Beispiel 8.3.1. Wir betrachten die Arbeit einer Telefonzentrale. Hier verstehen wir unter einem Signal einfach das Signal eines Abonnenten, der eine Telefonverbindung herstellen will. Die Bedingung I ist erfüllt, die Anzahlen der Telefongespräche in verschiedenen Zeitintervallen sind mit guter Näherung unabhängig. Die Bedingung II ist erfüllt, wenn wir nur unsere Beobachtungen immer zu derselben Tageszeit anstellen, so daß die Häufigkeit der angemeldeten Gespräche gleich bleibt. Auch die Bedingung III ist näherungsweise erfüllt.

Der betrachtete homogene Poissonsche Prozeß ist als Prozeß mit unabhängigen Zuwächsen, der der Gleichung $P(X_0 = 0) = 1$ genügt, ein homogener Markoffscher Prozeß. Die Übergangswahrscheinlichkeitsfunktion $p_{ij}(t)$ hat die Gestalt

$$p_{ij}(t) = P(X_{t_2} = j \mid X_{t_1} = i) = \frac{P(X_{t_1} = i,\ X_{t_2} - X_{t_1} = j - i)}{P(X_{t_1} = i)}$$

$$= P(X_{t_2} - X_{t_1} = j - i) = \frac{(\lambda t)^{j-i}}{(j-i)!} \exp(-\lambda t); \tag{8.3.15}$$

dabei ist $t = t_2 - t_1$ und $j \geqq i$. Offensichtlich ist $p_{ij}(t) = 0$ für $j < i$.

Aus der letzten Formel folgt, daß für einen homogenen Poissonschen Prozeß ein Ergodensatz gilt. Es gilt nämlich für beliebige i und j

$$\lim_{t\to\infty} p_{ij}(t) = 0. \tag{8.3.16}$$

Die Tatsache, daß alle Grenzwahrscheinlichkeiten gleich Null sind, erklärt sich daraus, daß alle Zustände vorübergehend sind (vgl. Aufgabe 7.6.5 und 7.6.9); denn von jedem Zustand i kann man zum Zustand $j > i$ übergehen, während man aber aus dem Zustand $j > i$ nicht mehr zum Zustand i zurückkehren kann.

Aus der Formel (8.3.15) ergeben sich leicht für $i = 0, 1, 2, \ldots$ die Beziehungen

$$q_i = \lim_{t\to 0} \frac{1 - p_{ii}(t)}{t} = \lim_{t\to 0} \frac{1 - \exp(-\lambda t)}{t} = \lambda, \tag{8.3.17}$$

$$q_{ij} = \lim_{t\to 0} \frac{p_{ij}(t)}{t} = \lim_{t\to 0} \frac{(\lambda t)^{j-i} \exp(-\lambda t)}{(j-i)!\,t}$$

$$= \begin{cases} \lambda & \text{für } j = i + 1, \\ 0 & \text{für } j > i + 1 \text{ und } j < i, \end{cases} \tag{8.3.18}$$

$$\sum_{\substack{j \\ j \neq i}} q_{ij} = q_i. \tag{8.3.19}$$

Die Ausdrücke q_i und q_{ij} heißen *Intensitäten* des stochastischen Prozesses. Wie man sieht, sind die Intensitäten eines homogenen Poissonschen Prozesses konstant (d. h. von t unabhängig).

Die Realisierungen eines homogenen Poissonschen Signalprozesses werden durch nichtabnehmende Treppenlinien dargestellt. Dem Auftreten eines Signals im Moment t entspricht ein Sprung der zufälligen Funktion um eine Einheit; dabei ist die Anzahl der Sprünge in jedem endlichen Intervall $[0, t)$ endlich. Abb. 8.3.1 stellt eine derartige Realisierung dar.

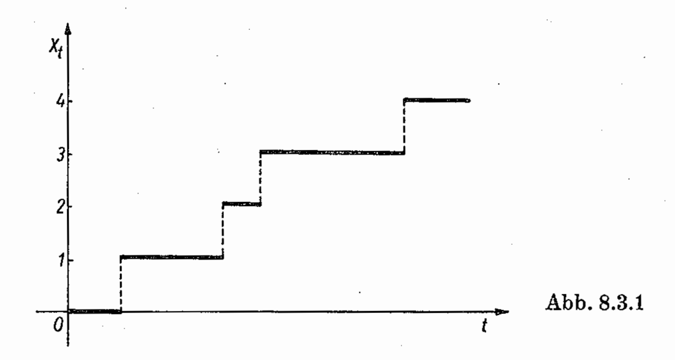


Abb. 8.3.1

Eigenschaften des homogenen Poissonschen Prozesses untersuchte Lévy [3].

Der Poissonsche Prozeß, seine verschiedenartigen Verallgemeinerungen und zahlreichen Anwendungen bildeten den Gegenstand vieler Arbeiten. Wir erinnern hier an die Arbeiten von Chintschin [2], [7], Florek, Marczewski und Ryll-Nardzewski [1], Marczewski [1], Prékopa [1], Jánossy, Rényi und Aczél [1], Rényi [1], [2], [6], Takács [2], [3], Fortet [3], Fisz und Urbanik [1]. Der gleichen Problematik ist ein ganzes Kapitel im Buch von Blanc-Lapierre und Fortet [1] gewidmet. (Siehe auch die Aufgaben 8.13.1 bis 8.13.8.).

Wir erinnern insbesondere an das kürzlich erzielte Ergebnis von Pyke [1], der eine Formel für die Verteilungsfunktion für die unteren und oberen Grenzen der Realisierung einer linearen Funktion $X_t - \alpha t$ eines Poissonschen Prozesses X_t in dem Intervall $[0, t_0)$ für ein beliebiges positives t_0 angegeben hat.

8.4. Der Furry-Yulesche Prozeß

Trotz der vielen Anwendungen, die der Poissonsche Prozeß gefunden hat, kann man ihn in vielen Problemen nur als erste Näherung ansehen. Um weitere und bessere Näherungen zu gewinnen, muß man viel allgemeinere Modelle betrachten. Diese Verallgemeinerungen können in zwei Richtungen erfolgen. Die erste Verallgemeinerung besteht darin, daß die Intensität λ, die durch die Formel (8.3.2) definiert ist, von dem Zustand, in dem sich der Prozeß befindet, abhängen kann. Bei der zweiten Verallgemeinerung betrachtet man Prozesse, die nicht nur das Auftreten gewisser Erscheinungen, sondern auch ihr Verschwinden berücksichtigen.

Jetzt beschäftigen wir uns näher mit der ersten Richtung, die zweite werden wir in 8.5 behandeln.

A. Es bezeichne X_t die Anzahl der nicht verschwindenden Elemente einer gewissen Menge zur Zeit t oder die Anzahl gewisser Ereignisse, die im Augenblick t oder früher eintraten. Dann können die Variablen X_t und die Zuwächse von X_t in einem beliebigen endlichen Intervall als Werte nur die ganzen Zahlen $j = 0, 1, 2, 3, \ldots$ annehmen. Der betrachtete Prozeß $\{X_t, 0 \leqq t < \infty\}$ genüge den Bedingungen I und II sowie der abgeänderten Bedingung III aus 8.3. Es soll nämlich jetzt die durch Formel (8.3.2) definierte Größe λ eine Funktion des Zustandes j sein, in dem sich der Prozeß im Moment t befindet, und wir wollen λ_j an Stelle von λ schreiben. Ein solcher Prozeß heißt *Geburtsprozeß*.

Die Gleichungen (8.3.2) und (8.3.3) nehmen hier für $j = 0, 1, 2, \ldots$ die Form

$$\lim_{\Delta t \to 0} \frac{p_{j,j+1}(\Delta t)}{\Delta t} = \lambda_j, \tag{8.4.1}$$

$$\lim_{\Delta t \to 0} \frac{1 - p_{jj}(\Delta t) - p_{j,j+1}(\Delta t)}{\Delta t} = 0 \tag{8.4.2}$$

an. Mit anderen Worten, für genügend kleine Δt ist die Wahrscheinlichkeit für das Erscheinen eines neuen Elements während des Zeitabschnitts $[t, t + \Delta t)$ gleich $\lambda_j \Delta t + o(\Delta t)$, wenn im Augenblick t die Anzahl der gerade vorhandenen Elemente j betrug, während die Wahrscheinlichkeit für die Geburt von Zwillingen oder Drillingen usw. (d. h. für wenigstens zwei Elemente) während derselben Zeit gleich $o(\Delta t)$ ist.

Es sei nun i eine gewisse fest gewählte natürliche Zahl. Wir nehmen an, daß $P(X_0 = i) = 1$ ist, und führen folgende Bezeichnung ein:

$$V_m(t) = P(X_t = i + m) \quad (m = 0, 1, 2, \ldots). \tag{8.4.3}$$

Wenn wir ebenso wie beim Beweis des Satzes 8.3.1 schließen, erhalten wir

$$V_0(t + \Delta t) = V_0(t)\, V_0(\Delta t),$$

$$V_m(t + \Delta t) = \sum_{k=0}^{m} V_{m-k}(t) V_k(\Delta t) \qquad (m = 1, 2, \ldots).$$

Die Formeln (8.4.1) und (8.4.2) ergeben

$$V_1(\Delta t) = \lambda_i \Delta t + o(\Delta t),$$

$$V_0(\Delta t) = 1 - \lambda_i \Delta t + o(\Delta t).$$

Aus den letzten vier Formeln erhält man das lineare Differentialgleichungssystem

$$V_0'(t) = -\lambda_i V_0(t), \tag{8.4.4}$$

$$V_m'(t) = -\lambda_{i+m} V_m(t) + \lambda_{i+m-1} V_{m-1}(t) \qquad (m = 1, 2, \ldots).$$

Aus der Voraussetzung $P(X_0 = i) = 1$ folgt, daß die Funktionen $V_m(t)$ die Anfangsbedingungen

$$V_m(0) = \begin{cases} 1 & \text{für} \quad m = 0, \\ 0 & \text{für} \quad m \neq 0 \end{cases} \tag{8.4.5}$$

erfüllen. Die erste Gleichung (8.4.4) ergibt also

$$V_0(t) = \exp(-\lambda_i t).$$

Die übrigen Funktionen $V_m(t)$ kann man sukzessive aus dem Gleichungssystem (8.4.4) bestimmen.

B. Wichtig ist der Spezialfall

$$\lambda_j = \lambda j \qquad (j = 1, 2, 3, \ldots). \tag{8.4.6}$$

Einen solchen Prozeß nennt man einen *Furry-Yuleschen Prozeß*. Furry [1] wandte diesen Prozeß bei der Untersuchung der kosmischen Strahlung an (vgl. Beispiel 8.4.1), während Yule [1] ihn in der Biologie (vgl. Beispiel 8.4.2) anwandte.

Die Gleichung (8.4.6) hat den folgenden Sinn: Die Wahrscheinlichkeit, daß ein schon vorhandenes Element während der Zeit Δt ein neues gebiert, ist näherungsweise gleich $\lambda \Delta t + o(\Delta t)$. Es seien nun in einem gewissen Moment schon j Elemente vorhanden, und es seien die Ereignisse, daß schon vorhandene Elemente neue gebären, voneinander unabhängig. Dann ist die Wahrscheinlichkeit, daß während der Zeit Δt ein neues Element geboren wird, näherungsweise gleich $\lambda j \Delta t + o(\Delta t)$, während die Wahrscheinlichkeit für die Geburt wenigstens zwei neuer Elemente gleich $o(\Delta t)$ ist.

Das Gleichungssystem (8.4.4) nimmt für einen Furry-Yuleschen Prozeß die Gestalt

$$V_0'(t) = -\lambda i V_0(t), \tag{8.4.7}$$

$$V_m'(t) = -\lambda(i + m) V_m(t) + \lambda(i + m - 1) V_{m-1}(t) \qquad (m = 1, 2, \ldots)$$

an. Das ist ein lineares Differentialgleichungssystem. Wenn wir die Gleichungen sukzessive auflösen und die Anfangsbedingungen (8.4.5) berücksichtigen, erhalten wir

$$V_m(t) = \binom{m + i - 1}{m} e^{-i\lambda t}(1 - e^{-\lambda t})^m \qquad (m = 0, 1, 2, \ldots). \tag{8.4.8}$$

Wir bemerken noch, daß für beliebige t die Gleichung

$$\sum_{m=0}^{\infty} V_m(t) = e^{-i\lambda t} \sum_{m=0}^{\infty} \binom{m+i-1}{m} (1-e^{-\lambda t})^m = 1 \tag{8.4.9}$$

gilt. Genügen also die λ_j den Gleichungen (8.4.6), dann ist mit der Wahrscheinlichkeit Eins die Anzahl der Elemente in einem beliebigen endlichen Zeitabschnitt endlich.

Hier kann man wie in 8.3 leicht zeigen, daß der betrachtete Prozeß ein homogener Markoffscher Prozeß ist. Für beliebige t_1 und t_2 $(0 \leqq t_1 < t_2 < \infty)$ und beliebige $m, j\,(m \leqq j)$ besteht die Gleichung

$$p_{mj}(t_1, t_2) = \binom{j-1}{j-m} e^{-m\lambda t}(1-e^{-\lambda t})^{j-m} = V_{j-m}(t) = p_{mj}(t)\,; \tag{8.4.10}$$

dabei ist $t = t_2 - t_1$.

Weiter gilt für diesen Prozeß der Ergodensatz; es ist nämlich

$$\lim_{t\to\infty} p_{mj}(t) = 0 \qquad (m = 1, 2, \ldots; j \geqq m). \tag{8.4.11}$$

Die Intensitäten werden für $m = 1, 2, 3, \ldots$ durch

$$q_m = \lim_{t\to 0} \frac{1 - p_{mm}(t)}{t} = m\lambda, \tag{8.4.12}$$

$$q_{mj} = \lim_{t\to 0} \frac{p_{mj}(t)}{t} = \begin{cases} m\lambda & \text{für } j = m+1, \\ 0 & \text{für } j \neq m, m+1 \end{cases} \tag{8.4.13}$$

gegeben, und es ist

$$\sum_{\substack{j \\ j\neq m}} q_{mj} = q_m. \tag{8.4.14}$$

Beispiel 8.4.1. Die kosmische Strahlung besteht hauptsächlich aus zwei Bestandteilen: aus der sogenannten harten und der weichen Strahlung. Die harte Strahlung setzt sich größtenteils aus Mesonen zusammen; ihren Namen verdankt sie der Tatsache, daß sie sogar durch meterdicke Bleiwände hindurchgeht. Die weiche Strahlung besteht aus Negatronen, Positronen (wir wollen sie Elektronen nennen) und aus Photonen; ihr Name stammt daher, daß schon eine 10 cm dicke Bleiwand diese Strahlung vollständig absorbiert. Die Theorie der harten Strahlung sowie die Deutung der erhaltenen experimentellen Ergebnisse ist noch nicht weit entwickelt. Anders ist es in der Theorie der weichen Strahlung. Arley [1] widmete den verschiedenen stochastischen Prozessen, die Anwendung in der Theorie der weichen Strahlung finden können, eine ausführliche Monographie. Dieser Monographie entnehmen wir das Material für das folgende Beispiel, bei dem zwei radioaktive Umwandlungen erfolgen können:

1. Ein Photon durcheilt in einem gewissen Medium den Weg Δt und unterliegt mit einer gewissen Wahrscheinlichkeit der Absorption, indem es zwei Elektronen emittiert.

2. Ein Elektron kann mit einer gewissen Wahrscheinlichkeit auf Kosten seiner Energie ein Photon emittieren.

Ein Elektron (oder Photon) kann also den Anfang einer Kette von aufeinanderfolgenden Umwandlungen bilden. Wenn nämlich die erste „Generation" aus einem Elektron besteht, so besteht die zweite aus einem Photon; dieses wiederum kann das Entstehen zweier Elektronen in der dritten „Generation" bewirken usw. Es entsteht also ein Elektronen- oder Photonenstrom. Wir sprechen hier von einer „Kaskade" oder einem „Kaskadenprozeß". Es sind nur experimentelle Beobachtungen der Elektronen möglich, es kann also nur jede zweite „Generation" beobachtet werden.

Es sei X_t die Anzahl der Elektronen in der Tiefe t (die „Zeit" t bedeutet hier also Tiefe). FURRY [1] vernachlässigte die mittlere Photonengeneration und nahm an, daß ein Elektron, das den Weg Δt durchlaufen hat, sich mit der Wahrscheinlichkeit $\lambda \Delta t + o(\Delta t)$ in zwei Elektronen umwandeln kann und daß jedes Elektron unabhängig von den übrigen Elektronen dieselben Möglichkeiten der Umwandlung hat, so daß die Gleichung (8.4.6) erfüllt ist. Diese Annahmen bedeuten, daß der Kaskadenprozeß dem Gleichungssystem (8.4.7) genügt. Das Furrysche Modell ist eine bessere Näherung im Vergleich zur Arbeit von BHABHA und HEITLER [1], die annahmen, daß der Kaskadenprozeß ein Poissonscher Prozeß ist. Aber auch das Furrysche Modell ist noch ungenügend, da es noch nicht die Möglichkeit berücksichtigt, daß Elektronen infolge eines Energieverlustes bei der Emission von Photonen verschwinden können. ARLEY [1] betrachtete ein Modell, das den eben erwähnten Ansprüchen genügt und gut mit den experimentellen Ergebnissen übereinstimmt. Einzelheiten findet der Leser in der zitierten Arleyschen Arbeit.

Beispiel 8.4.2. Den Geburtsprozeß wandte YULE [1] auf eine biologische Gesamtheit an, in der sich die Elemente durch Zweiteilung fortpflanzen. Wenn wir annehmen, daß die einzelnen Elemente der Gesamtheit keinerlei Einfluß aufeinander ausüben, d. h., daß die Teilungen unabhängig voneinander erfolgen, so können wir das oben betrachtete Modell anwenden. Zweckentsprechend ist ein Modell, das auch die Todesfälle berücksichtigt. Ein solches Modell hat FELLER [3] ausgearbeitet.

C. Wir wollen den Mittelwert $E(X_t)$ und die Dispersion $D^2(X_t)$ des Geburtenprozesses berechnen. Dazu benutzen wir eine von ARLEY [1] stammende Methode. Wir führen folgende Bezeichnungen ein:

$$E(X_t) = E(t), \quad E(X_t^2) = E_2(t), \quad D^2(X_t) = D^2(t).$$

Es ist

$$E(t) = \sum_{m=0}^{\infty} (i+m) V_m(t).$$

Aus den Formeln (8.4.7) erhalten wir[1])

$$E'(t) = \sum_{m=0}^{\infty} (i+m) V'_m(t) = \lambda \sum_{m=0}^{\infty} (i+m) V_m(t) = \lambda E(t),$$

[1]) Man zeigt leicht, daß die Differentiation unter dem Summenzeichen gestattet ist.

also ergibt sich bei Berücksichtigung der Anfangsbedingungen $E(0) = i$

$$E(t) = i e^{\lambda t}. \tag{8.4.15}$$

Weiter ist

$$E_2(t) = \sum_{m=0}^{\infty} (i + m)^2 V_m(t);$$

daraus folgt[1])

$$\big(E_2(t)\big)' = \sum_{m=0}^{\infty} (i + m)^2 V'_m(t) = \lambda \sum_{m=0}^{\infty} (i + m)(2i + 2m + 1) V_m(t)$$

$$= 2\lambda E_2(t) + \lambda E(t).$$

Wenn wir die Anfangsbedingungen $E_2(0) = i^2$ berücksichtigen, erhalten wir

$$E_2(t) = e^{2\lambda t}[i^2 + i(1 - e^{-\lambda t})] \tag{8.4.15'}$$

als Lösung dieser linearen Differentialgleichung. Daraus und aus (8.4.15) ergibt sich

$$D^2(t) = E_2(t) - [E(t)]^2 = i(e^{2\lambda t} - e^{\lambda t}). \tag{8.4.16}$$

D. Oben erwähnten wir, daß beim Furry-Yuleschen Prozeß für beliebige t die Gleichung (8.4.9) gilt. Die Funktionen $V_m(t)$, die das Differentialgleichungssystem (8.4.4) lösen, brauchen aber nicht für beliebige λ_j die Gleichung (8.4.9) zu erfüllen; es kann nämlich

$$\sum_{m=0}^{\infty} V_m(t) < 1 \tag{8.4.17}$$

sein. Das bedeutet, daß während eines gewissen endlichen Zeitabschnitts mit positiver Wahrscheinlichkeit unendlich viele Elemente auftreten können, und das ist dann der Fall, wenn λ_j für $j \to \infty$ schnell gegen ∞ strebt. Dieser Sachverhalt folgt aus nachstehendem Satz, der ein Spezialfall des Feller-Lundbergschen Satzes ist.[2])

Satz 8.4.1. *Die Beziehung*

$$\sum_{m=0}^{\infty} \frac{1}{\lambda_{i+m}} = \infty \tag{8.4.18}$$

---

[1]) Man zeigt leicht, daß die Differentiation unter dem Summenzeichen gestattet ist.

[2]) Der Feller-Lundbergsche Satz betrifft auch inhomogene Geburtsprozesse (vgl. Feller [4] und Lundberg [1]).

ist eine notwendige und hinreichende Bedingung dafür, daß die Lösungsfunktionen $V_m(t)$ *von* (8.4.4), *die die Bedingungen* (8.4.5) *erfüllen, der Gleichung* (8.4.9) *genügen.*

Beweis. Wir setzen

$$W_m(t) = \sum_{r=0}^{m} V_r(t) \qquad (m = 0, 1, 2, \ldots)$$

und erhalten damit aus (8.4.4)

$$W'_m(t) = \sum_{r=0}^{m} V'_r(t) = -\lambda_{i+m} V_m(t).$$

Wenn wir beide Seiten der letzten Gleichung integrieren und die Anfangsbedingungen berücksichtigen, ergibt sich

$$1 - W_m(t) = \lambda_{i+m} \int_0^t V_m(t)\, dt. \tag{8.4.19}$$

Da die Funktionen $V_r(t)$ nicht negativ sind, ist $1 - W_m(t)$ bei festem t eine nicht wachsende Funktion des Arguments m. Es existiert also auch der Grenzwert

$$U(t) = \lim_{m\to\infty} [1 - W_m(t)],$$

woraus sich

$$\lambda_{i+m} \int_0^t V_m(t)\, dt \geqq U(t)$$

ergibt. Es ist weiter

$$\int_0^t W_m(t)\, dt = \sum_{r=0}^{m} \int_0^t V_r(t)\, dt \geqq U(t) \sum_{r=0}^{m} \frac{1}{\lambda_{i+r}}. \tag{8.4.20}$$

Es sei nun (8.4.18) erfüllt. Die linke Seite von (8.4.20) ist nicht größer als t, da $W_m(t) < 1$ auf Grund der Beziehung (8.4.17) ist. Dagegen wächst die rechte Seite von (8.4.20) für $m \to \infty$ gegen Unendlich, wenn $U(t) > 0$ ist. Damit die Ungleichung (8.4.20) gilt, muß $U(t) \equiv 0$, also $\lim_{m\to\infty} W_m(t) = 1$ sein, und das bedeutet, daß die Gleichung (8.4.9) erfüllt sein muß.

Nun sei die Gleichung (8.4.9) für jedes t erfüllt. Aus (8.4.19) folgen für jedes r die Ungleichungen

$$\int_0^t V_r(t)\, dt \leqq \frac{1}{\lambda_{i+r}},$$

und somit gilt für alle m und t die Ungleichung

$$\int_0^t \sum_{r=0}^m V_r(t)\,dt \leqq \sum_{r=0}^m \frac{1}{\lambda_{i+r}}.$$

Nun vollziehen wir den Grenzübergang auf beiden Seiten dieser Ungleichung für $m \to \infty$. Nach (8.4.9) erhalten wir dann auf der linken Seite t. Da dies für jedes t gilt, muß die auf der rechten Seite gewonnene Reihe divergent und damit die Beziehung (8.4.18) erfüllt sein. Damit ist der Satz bewiesen.

Der Satz 8.4.1 ist ein Spezialfall des allgemeinen Satzes von Feller [4], der die Bedingungen dafür angibt, daß die Gleichung (8.4.9) für einen homogenen Markoffschen Prozeß mit abzählbar vielen Zuständen gilt. Lundberg [1] hat den Satz 8.4.1 auf einen inhomogenen Geburtsprozeß verallgemeinert, d. h. auf den Fall, daß die Koeffizienten λ_j Funktionen von t sind. Mit einer weiteren Verallgemeinerung des Satzes 8.4.1 werden wir uns in 8.5.A beschäftigen.

Wie wir sehen, erfüllen die Intensitäten eines Furry-Yuleschen Prozesses die Beziehung (8.4.18).

Ist die Ungleichung (8.4.17) erfüllt, so kann sich der Prozeß während eines gewissen Zeitintervalls im Zustand „unendlich" befinden.

Doob [3] hat gezeigt, daß man auch diesen Spezialfall wahrscheinlichkeitstheoretisch deuten kann, indem man allgemein sagt, daß der Prozeß von einem Zustand zum anderen in unendlich vielen Schritten übergeht. Dieses Problem ist sehr subtil. Den interessierten Leser verweisen wir auf den schon zitierten Artikel von Doob.

8.5. Geburts- und Todesprozesse

Wir erhalten eine weitere Verallgemeinerung der in 8.4 behandelten Prozesse, wenn wir annehmen, daß bei dem Prozeß Elemente nicht nur hinzukommen, sondern auch verschwinden können. In dem betrachteten physikalischen Beispiel über die Radioaktivität wird ein Modell, das auch die Absorption von Elementarteilchen berücksichtigt, der Wirklichkeit näherkommen als ein Modell, das nur die Entstehung der Teilchen berücksichtigt. Auch bei der Arbeit einer Telefonzentrale wird ein Modell, das nicht nur die Anzahl der begonnenen Gespräche berücksichtigt, wie das schon beim Signalprozeß der Fall war, sondern auch die Anzahl der während eines gewissen Zeitabschnitts endenden Gespräche zählt, eine bessere mathematische Idealisierung der wirklichen Bedingungen sein, unter denen eine Telefonzentrale arbeitet. Derartig verallgemeinerte Prozesse nennt man *Geburts- und Todesprozesse.* Weitgehende Verallgemeinerungen wollen wir nicht behandeln, sondern uns auf die Betrachtung einer möglichst einfachen Verallgemeinerung in dieser Richtung beschränken, die das Wesentliche des Problems klar zum Ausdruck bringt.

A. Wir betrachten wieder einen Signalprozeß $\{X_t, 0 \leqq t < \infty\}$, der den Bedingungen I bis III von 8.3 genügt und die Intensität λ besitzt. Wir wollen die „Lebensdauer" T eines Signals betrachten, die eine Zufallsvariable ist. Diese möge eine Exponentialverteilung haben. Ihre Dichtefunktion $g(t)$ hat dann die Form

$$g(t) = \begin{cases} 0 & \text{für} \quad t \leqq 0, \\ \mu e^{-\mu t} & \text{für} \quad t > 0, \end{cases}$$

wobei $\mu > 0$ ist.

Die Voraussetzung, daß die Lebensdauer eines Signals eine Exponentialverteilung haben soll, ist in vielen Anwendungen eine gute Näherung und hat auch theoretisch große Bedeutung. In der Tat erhalten wir unter diesen Voraussetzungen

$$\lim_{\Delta t \to 0} \frac{P(T < t + \Delta t \mid T \geqq t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(t \leqq T < t + \Delta t)}{\Delta t P(T \geqq t)}$$

$$= \frac{g(t)}{P(T \geqq t)} = \mu.$$

Also ist die Wahrscheinlichkeit dafür, daß das Signal im Zeitintervall $[t, t + \Delta t)$ erlischt, nachdem es schon t Zeiteinheiten gedauert hat, für kleine Δt gleich $\mu \Delta t + o(\Delta t)$, und diese Wahrscheinlichkeit hängt nicht von t ab, d. h., sie hängt nicht davon ab, wie lange das Signal schon dauerte. Unter diesen Bedingungen ist das Erscheinen und Erlöschen der Signale ein Markoffscher Prozeß.

Wir setzen außerdem noch voraus, daß die Lebensdauern der einzelnen Signale voneinander unabhängig sind und auch unabhängig von der Anzahl der beginnenden Signale sein sollen. Weiter sei die Anzahl der Signale, die in einem gewissen Intervall erlöschen, nur von der Länge dieses Intervalls abhängig, nicht aber von den Endpunkten dieses Intervalls. Hört man also i Signale im Moment t, so ist die Wahrscheinlichkeit dafür, daß ein Signal im Zeitintervall $[t, t + \Delta t)$ erlischt, gleich $\mu i \Delta t + o(\Delta t)$, während die Wahrscheinlichkeit für das Erlöschen von wenigstens zwei Signalen von der Ordnung $o(\Delta t)$ sei. Ebenso sei die Wahrscheinlichkeit dafür, daß in der Zeit Δt ein neues Signal auftaucht und ein schon vorhandenes erlischt, ebenfalls von der Ordnung $o(\Delta t)$.

Wir bezeichnen mit Y_t die Anzahl der im Moment t hörbaren Signale und betrachten den Prozeß $\{Y_t, 0 \leqq t < \infty\}$. Der Prozeß soll im Zustand j sein, wenn j Signale im Augenblick t hörbar sind. Wir wollen die totale Wahrscheinlichkeit $c_j(t)$ dafür berechnen, daß sich im Augenblick t der Prozeß im Zustand j befindet:

$$c_j(t) = P(Y_t = j) \qquad (j = 0, 1, 2, \ldots).$$

Es ist

$$c_0(t + \Delta t) = c_0(t)[1 - \lambda\Delta t + o(\Delta t)] + c_1(t)[\mu\Delta t + o(\Delta t)] + o(\Delta t), \tag{8.5.1}$$

$$\begin{aligned} c_j(t + \Delta t) = c_j(t)&[1 - \lambda\Delta t + o(\Delta t)][1 - \mu j\Delta t + o(\Delta t)] \\ &+ c_{j-1}(t)[\lambda\Delta t + o(\Delta t)] \\ &+ c_{j+1}(t)[\mu(j + 1)\Delta t + o(\Delta t)] + o(\Delta t) \qquad (j = 1, 2, \ldots). \end{aligned} \tag{8.5.2}$$

In der Tat, der Prozeß kann im Augenblick $t + \Delta t$ nur dann im Zustand Null sein, wenn er entweder im Moment t in diesem Zustand war und während des Intervalls $[t, t + \Delta t)$ kein neues Signal hinzugekommen ist oder wenn der Prozeß im Augenblick t im Zustand Eins war und im Intervall $[t, \Delta t + t)$ dieses Signal erloschen ist. Damit haben wir Gleichung (8.5.1) begründet.

Weiter kann der Prozeß im Augenblick $t + \Delta t$ im Zustand j $(j \geqq 1)$ sein, wenn er entweder schon im Augenblick t im Zustand j war, in dem Intervall $[t, t + \Delta t)$ kein neues Signal hinzugekommen und auch kein altes erloschen ist oder wenn der Prozeß im Augenblick t im Zustand $j - 1$ war und in dem Intervall $[t, t + \Delta t)$ ein neues Signal hinzugekommen ist oder wenn der Prozeß im Augenblick t im Zustand $j + 1$ war und in dem Intervall $[t, t + \Delta t)$ ein vorhandenes Signal erloschen ist. Die Wahrscheinlichkeit aller anderen Möglichkeiten ist von der Ordnung $o(\Delta t)$. Damit haben wir die Gleichung (8.5.2) hergeleitet.

Aus diesen Gleichungen erhalten wir die Differentialgleichungen

$$\begin{aligned} c_0'(t) &= -\lambda c_0(t) + \mu c_1(t), \\ c_j'(t) &= \lambda c_{j-1}(t) - (\lambda + j\mu)c_j(t) + \mu(j + 1)c_{j+1}(t) \qquad (j = 1, 2, \ldots). \end{aligned} \tag{8.5.3}$$

Wir haben also wieder ein unendliches System linearer Differentialgleichungen erhalten, die aber diesmal keine rekursiven Beziehungen sind. Dieses Gleichungssystem wollen wir mit Hilfe von erzeugenden Funktionen lösen. Dazu führen wir die Bezeichnung

$$f(t, s) = \sum_{j=0}^{\infty} c_j(t) s^j \tag{8.5.4}$$

ein und erhalten

$$\begin{aligned} \frac{\partial f(t, s)}{\partial t} = \sum_{j=0}^{\infty} \frac{dc_j(t)}{dt} s^j &= -\lambda c_0(t)(1 - s) + \mu c_1(t)(1 - s) \\ &\quad - \lambda c_1(t)(1 - s)s + 2\mu c_2(t)s(1 - s) + \cdots \\ &= -\lambda(1 - s)\sum_{j=0}^{\infty} c_j(t)s^j + \mu(1 - s)\sum_{j=1}^{\infty} j c_j(t) s^{j-1} \\ &= -\lambda(1 - s)f(t, s) + \mu(1 - s)\frac{\partial f(t, s)}{\partial s}. \end{aligned}$$

Die erzeugende Funktion $f(t, s)$ erfüllt also die lineare partielle Differentialgleichung erster Ordnung

$$\frac{\partial f}{\partial t} - \mu(1 - s)\frac{\partial f}{\partial s} = -\lambda(1 - s) f. \tag{8.5.5}$$

Wir wollen annehmen, daß im Moment $t = 0$

$$c_i(0) = 1, \quad c_j(0) = 0 \quad (j \neq i)$$

ist. Die Funktion $f(t, s)$ erfüllt also die Anfangsbedingung

$$f(0, s) = s^i. \tag{8.5.6}$$

Durch die Beziehung $V(s, t, f) = 0$ definieren wir eine neue Funktion V. Aus (8.5.5) ergibt sich mit Hilfe von

$$\frac{\partial f}{\partial s} = -\frac{\partial V}{\partial s} : \frac{\partial V}{\partial f}, \qquad \frac{\partial f}{\partial t} = -\frac{\partial V}{\partial t} : \frac{\partial V}{\partial f}$$

für V die homogene lineare partielle Differentialgleichung erster Ordnung

$$\frac{\partial V}{\partial t} - \mu(1 - s)\frac{\partial V}{\partial s} - \lambda(1 - s) f \frac{\partial V}{\partial f} = 0. \tag{8.5.7}$$

Die Gleichung (8.5.7) ist dem gewöhnlichen linearen Differentialgleichungssystem

$$dt = -\frac{ds}{\mu(1 - s)},$$

$$ds = \frac{\mu}{\lambda}\frac{df}{f}$$

äquivalent, dessen Lösung die Form

$$\mu t - \log(1 - s) = a_1,$$

$$\lambda s - \mu \log f = a_2$$

hat. Dann besitzt die Lösung der Gleichung (8.5.7) die Gestalt

$$\lambda s - \mu \log f = \psi[\mu t - \log(1 - s)];$$

dabei ist ψ eine beliebige Funktion. Daraus erhalten wir

$$f = \exp\left\{\frac{1}{\mu}\left[\lambda s - \psi\left(\mu t - \log(1 - s)\right)\right]\right\}. \tag{8.5.8}$$

Aus der Anfangsbedingung (8.5.6) kann man die Funktion ψ bestimmen. Wir erhalten nämlich

$$\mu i \log s = \lambda s - \psi\left(\log \frac{1}{1-s}\right).$$

Wenn wir $y = \log \dfrac{1}{1-s}$, also $s = 1 - e^{-y}$ setzen, ergibt sich

$$\psi(y) = \lambda(1 - e^{-y}) - i\mu \log(1 - e^{-y}),$$

also

$$\psi[\mu t - \log(1-s)] = \lambda[1 - (1-s)e^{-\mu t}] - \mu i \log[1 - (1-s)e^{-\mu t}].$$

Schließlich finden wir

$$f(t, s) = [1 - (1-s)e^{-\mu t}]^i \exp\left[-\frac{\lambda}{\mu}(1-s)(1 - e^{-\mu t})\right]. \tag{8.5.9}$$

Der zweite Faktor auf der rechten Seite von (8.5.9) stellt (vgl. 4.7) die erzeugende Funktion einer Zufallsvariablen dar, die eine Poissonverteilung besitzt und deren Mittelwert (also auch deren Dispersion) gleich

$$\frac{\lambda}{\mu}(1 - e^{-\mu t})$$

ist; der erste Faktor ist die erzeugende Funktion einer Binomialverteilung, bei der die Wahrscheinlichkeit eines Erfolges gleich $e^{-\mu t}$ ist. Der Erfolg besteht hier darin, daß ein Signal mindestens während der Zeit t andauert. Also ist Y_t für jedes t die Summe zweier unabhängiger Zufallsvariabler, von denen die eine eine Poissonverteilung besitzt und die andere binomial verteilt ist. Daraus folgt

$$c_j(t) = \exp\left[-\frac{\lambda}{\mu}(1 - e^{-\mu t})\right] \sum_{k=0}^{\min(i,j)} \binom{i}{k} \left(\frac{\lambda}{\mu}\right)^{j-k} \cdot \frac{e^{-\mu t k}(1 - e^{-\mu t})^{i+j-2k}}{(j-k)!}$$

$$(j = 0, 1, 2, \ldots), \tag{8.5.10}$$

$$E(Y_t) = \frac{\lambda}{\mu}(1 - e^{-\mu t}) + i e^{-\mu t},$$

$$D^2(Y_t) = (1 - e^{-\mu t})\left(\frac{\lambda}{\mu} + i e^{-\mu t}\right).$$

Diese Formel stammt von Palm [1].

In dem Spezialfall $i = 0$ hat Y_t eine Poissonsche Verteilung, und es ist

$$c_j(t) = \exp\left[-\frac{\lambda}{\mu}(1 - e^{-\mu t})\right] \frac{\left[\frac{\lambda}{\mu}(1 - e^{-\mu t})\right]^j}{j!} \qquad (j = 0, 1, 2, \ldots). \tag{8.5.11}$$

Es ist

$$c_j(t) = \sum_l c_l(0) p_{lj}(t).$$

Aus der Anfangsbedingung $c_i(0) = 1$ folgen die Gleichungen

$$c_j(t) = p_{ij}(t) \qquad (j = 0, 1, 2, \ldots).$$

Da aus (8.5.10)

$$\lim_{t\to 0} c_j(t) = \begin{cases} 1 & \text{für } j = i, \\ 0 & \text{für } j \neq i \end{cases}$$

folgt, erhalten wir bei Berücksichtigung der Formeln (8.5.3) die Intensitäten

$$q_i = \lim_{t\to 0} \frac{1 - p_{ii}(t)}{t} = \lim_{t\to 0} \frac{1 - c_i(t)}{t} = -c_i'(0) = \lambda + i\mu, \tag{8.5.12}$$

$$q_{ij} = \lim_{t\to 0} \frac{p_{ij}(t)}{t} = c_j'(0) = \begin{cases} i\mu & \text{für } j = i - 1, \\ \lambda & \text{für } j = i + 1, \\ 0 & \text{für } j \neq i - 1, i, i + 1. \end{cases} \tag{8.5.13}$$

Es läßt sich leicht zeigen, daß für jedes t die Ungleichung

$$\sum_{j=0}^{\infty} c_j(t) = 1$$

gilt.

Wenn jedoch die Intensität λ für das Auftreten eines Signals eine Funktion der Anzahl der aktuell tätigen Signale ist und damit die Form λ_j hat, während die Intensität für das Verschwinden eines unter den i aktuell tätigen Signalen von der Form $\mu_i \neq \mu i$ ist, so kann die Ungleichung

$$\sum_{j=0}^{\infty} c_j(t) < 1$$

gelten.

Notwendige und hinreichende Bedingungen dafür, daß in der letzten Reihe für jedes t das Gleichheitszeichen gilt, wurden von Dobruschin [1] und von

Reuter und Ledermann [1] angegeben, die damit den Satz 8.4.1 verallgemeinert haben.

Aus (8.5.10) erkennt man, daß die Verteilung von Y_t für $t \to \infty$ gegen die Poissonsche Verteilung mit dem Mittelwert $\frac{\lambda}{\mu}$ konvergiert (unabhängig vom Anfangszustand). Also ist

$$c_j = \lim_{t\to\infty} c_j(t) = \frac{\left(\frac{\lambda}{\mu}\right)^j}{j!} \exp\left(-\frac{\lambda}{\mu}\right) \qquad (j = 0, 1, 2, \ldots). \tag{8.5.14}$$

Die Frage nach der Existenz der Grenzwerte c_j hat schon Erlang [1] in seinen grundlegenden Arbeiten über die Anwendung der Wahrscheinlichkeitsrechnung im Fernmeldewesen gestellt.

Beispiel 8.5.1. Wir betrachten die Tätigkeit einer Telefonzentrale, die soviel Leitungen besitzt, daß jedes in einem beliebigen Moment auftretende Signal eine freie Leitung vorfindet; mit anderen Worten, die Zentrale besitze unendlich viele Leitungen. Wenn wir mit Y_t die Anzahl der im Augenblick t belegten Leitungen bezeichnen, so sehen wir, daß Y_t ein Prozeß mit abzählbar vielen Zuständen ist. Die oben gefundene Formel (8.5.10) erlaubt uns, für jedes t die totale Wahrscheinlichkeit $c_j(t) = P(Y_t = j)$ zu berechnen, während die Formel (8.5.14) den Grenzwert c_j dieser Wahrscheinlichkeit für $t \to \infty$ angibt.

Beispiel 8.5.2. Wir betrachten jetzt die Tätigkeit einer Telefonzentrale mit endlich vielen Leitungen. Diese Telefonzentrale sei so beschaffen, daß ein Signal, das in einem Augenblick, in dem alle Leitungen belegt sind, auftaucht, nicht auf eine freie Leitung wartet, sondern wegfällt. Wir sagen, daß die Zentrale *mit Verlust* arbeitet. Das wichtigste Problem, das vor dem Bau einer solchen Zentrale zu lösen ist, ist die Bestimmung der Wahrscheinlichkeit dafür, daß im Augenblick t alle Leitungen belegt sind. Dieses Problem wollen wir nun behandeln; dabei beschränken wir uns auf die Bestimmung der Grenzwerte der gesuchten Wahrscheinlichkeit für $t \to \infty$.

B. Es sei Y_t wie bisher die Anzahl der im Augenblick t hörbaren Signale. Wir wollen alle bisherigen Voraussetzungen beibehalten mit dem einzigen Unterschied: Die Anzahl der Zustände dieses Prozesses sei endlich $(j = 0, 1, \ldots, m)$, also $c_j(t) = 0$ für $j > m$. Die Gleichungen (8.5.1) und (8.5.2) bleiben für $j = 1, \ldots, m-1$ richtig, während sich für $j = m$ die Gleichung

$$c_m(t + \Delta t) = c_m(t)[1 - \mu m \Delta t + o(\Delta t)] + c_{m-1}(t)[\lambda \Delta t + o(\Delta t)] + o(\Delta t)$$

ergibt. Wir erhalten also das Differentialgleichungssystem

$$c_0'(t) = -\lambda c_0(t) + \mu c_1(t), \tag{8.5.15}$$

$$c_j'(t) = \lambda c_{j-1}(t) - (\lambda + j\mu) c_j(t) + \mu(j+1) c_{j+1}(t) \quad (j = 1, \ldots, m-1),$$

$$c_m'(t) = \lambda c_{m-1}(t) - \mu m c_m(t)$$

und stellen uns die Aufgabe, den Grenzwert c_j der Funktion $c_j(t)$ für $t \to \infty$ zu finden. Dieser Grenzwert ist, wie wir gleich sehen werden, unabhängig von der Anfangsverteilung $c_j(0)$ $(j = 0, 1, \ldots, m)$. In der Tat, die Voraussetzungen des Satzes 8.2.1 sind erfüllt, d. h., es gibt ein solches t^*, daß die Übergangswahrscheinlichkeiten $p_{ij}(t^*)$ $(i, j = 0, 1, \ldots, m)$ positiv sind. Es gilt also der Ergodensatz, und es existieren die Wahrscheinlichkeiten $p_j = \lim_{t \to \infty} p_{ij}(t)$ für $j = 0, 1, \ldots, m$. Aus dem Satz 8.2.2 folgt dann die Existenz der Grenzwerte $c_j = \lim_{t \to \infty} c_j(t) = p_j$, die, wie wir wissen, von der Anfangsverteilung unabhängig sind. Wir haben also gezeigt, daß die rechten Seiten der Gleichungen (8.5.15) für $t \to \infty$ einem Grenzwert zustreben. Dann existieren auch die Grenzwerte der Funktionen $c_j'(t)$. Diese Ableitungen können aber nur gegen Null streben, da sonst $c_j(t)$ gegen Unendlich gehen würde. Also erhalten wir für die Bestimmung der Konstanten c_j das Gleichungssystem

$$\begin{aligned} &-\lambda c_0 + \mu c_1 = 0, \\ &\lambda c_{j-1} - (\lambda + j\mu) c_j + \mu(j+1) c_{j+1} = 0 \qquad (j = 1, \ldots, m-1), \\ &\lambda c_{m-1} - \mu m c_m = 0. \end{aligned} \tag{8.5.16}$$

Aus diesen Gleichungen ergibt sich

$$c_j = \frac{1}{j!} \left(\frac{\lambda}{\mu}\right)^j c_0 \qquad (j = 1, \ldots, m).$$

Berücksichtigen wir noch, daß $c_0 + \cdots + c_m = 1$ ist, so erhalten wir

$$c_0 = \frac{1}{\sum_{k=0}^{m} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k}$$

und schließlich

$$c_j = \frac{\frac{1}{j!} \left(\frac{\lambda}{\mu}\right)^j}{\sum_{k=0}^{m} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k} \qquad (j = 0, 1, \ldots, m). \tag{8.5.17}$$

Diese Formeln erhielt bereits ERLANG [1].

Beispiel 8.5.3. Tabelle 8.5.1 enthält die Werte c_j, die nach der Formel (8.5.17) berechnet wurden, für $m = 5$ und $m = 10$, wobei $\lambda = \mu$ und $\lambda = 2\mu$ gesetzt wurde. Aus dieser Tabelle liest man ab: Für $\lambda = 2\mu$ und fünf Zuleitungen ist die Wahrscheinlichkeit, daß alle Leitungen belegt sind, ziemlich groß. Sie beträgt nämlich ungefähr 0,04, d. h., ungefähr jeder fünfund-

zwanzigste Anruf geht verloren. Dagegen genügen 10 Leitungen für dasselbe Verhältnis $\lambda = 2\mu$ vollständig. Für $\lambda = \mu$ ist die Wahrscheinlichkeit, daß alle Leitungen belegt sind, schon für $m = 5$ sehr klein.

Tabelle 8.5.1. *Werte von c_j, berechnet nach (8.5.17)*

j	$m = 5$		$m = 10$	
	$\lambda = \mu$	$\lambda = 2\mu$	$\lambda = \mu$	$\lambda = 2\mu$
0	0,36810	0,13761	0,36788	0,13534
1	0,36810	0,27523	0,36788	0,27067
2	0,18405	0,27523	0,18394	0,27067
3	0,06135	0,18349	0,06131	0,18045
4	0,01534	0,09174	0,01533	0,09022
5	0,00307	0,03670	0,00307	0,03609
6			0,00051	0,01203
7			0,00007	0,00344
8			0,00001	0,00086
9			0,00000..	0,00019
10			0,00000..	0,00004

C. Wir betrachten jetzt einen anderen Typ einer Telefonzentrale mit endlich vielen Leitungen, nämlich eine *Zentrale mit Warteplätzen*. Diese Zentrale ist so konstruiert, daß die ankommenden Signale eine Warteschlange bilden, falls alle Leitungen besetzt sind, und so lange warten, bis eine Leitung frei wird. Für die Arbeit einer Zentrale dieses Typs sind zwei Probleme wesentlich:

1. Bestimmung der Wahrscheinlichkeitsfunktion für die Anzahl derjenigen Signale, die warten müssen;
2. Bestimmung der Verteilungsfunktion der Wartezeit.

Mit dem ersten Problem hat sich Kolmogoroff [4] beschäftigt. Wir wollen hier die Lösung des Problems skizzieren. Dabei beschränken wir uns darauf, die Grenzwerte der gesuchten Wahrscheinlichkeiten für $t \to \infty$ zu bestimmen.

Mit dem zweiten Problem hat sich schon Erlang [1] beschäftigt. Weiter unten bringen wir unter gewissen einfachen Voraussetzungen die Lösung des zweiten Problems.

Diese Problematik gibt es nicht nur im Telefonwesen, sondern auch in vielen anderen Gebieten der Technik und des täglichen Lebens. So kann man z. B. die Fahrkartenschalter eines Bahnhofs als „Zentrale" ansehen, ein „Signal" ist hier das Erscheinen eines Passagiers am Schalterfenster. Der Prozeß ist im Zustand j, wenn j Passagiere entweder eine Fahrkarte kaufen oder Schlange stehen. Es ist klar, daß die Eisenbahndirektion sowohl über den Andrang der Reisenden als auch über die Schnelligkeit der Abfertigung Erkundigungen am Schalter einziehen muß, um auf eine vernünftige Weise die Anzahl der nötigen Schalter zu regeln. Solche Prozesse heißen *Massenbedienungsprozesse*.

Es bezeichne jetzt Y_t die Anzahl der Signale, die im Augenblick t entweder andauern oder auf das Freiwerden einer Leitung warten. Wir nehmen wieder an, daß der Strom der aufleuchtenden Signale den Bedingungen I bis III von 8.3 genügt und die Intensität λ besitzt und daß die Signaldauer eine Exponentialverteilung mit dem Parameter μ hat. Ebenfalls wollen wir wieder die übrigen Voraussetzungen machen, die zum Gleichungssystem (8.5.3) führten. Es sei

$$c_j(t) = P(Y_t = j),$$

und die Anzahl der Leitungen sei gleich m. Die Gleichungen (8.5.1) und (8.5.2) bleiben für $0 < j < m$ unverändert, da sich in dem Fall wenigstens eine freie Leitung findet und das auftauchende Signal nicht zu warten braucht. Dagegen ist für $j \geqq m$

$$c_j(t + \Delta t) = c_j(t)[1 - \mu m \Delta t + o(\Delta t)][1 - \lambda \Delta t + o(\Delta t)]$$
$$+ c_{j-1}(t)[\lambda \Delta t + o(\Delta t)] + c_{j+1}(t)[\mu m \Delta t + o(\Delta t)] + o(\Delta t).$$

Die letzte Gleichung unterscheidet sich von (8.5.2) darin, daß auf der rechten Seite μm an Stelle von μj und $\mu(j+1)$ steht, weil die Anzahl der tätigen Signale nicht größer als m sein kann. Schließlich erhalten wir das Differentialgleichungssystem

$$c_0'(t) = -\lambda c_0(t) + \mu c_1(t), \tag{8.5.18}$$
$$c_j'(t) = \lambda c_{j-1}(t) - (\lambda + j\mu) c_j(t) + \mu(j+1) c_{j+1}(t) \quad (j = 1, 2, \ldots, m-1),$$
$$c_j'(t) = \lambda c_{j-1}(t) - (\lambda + m\mu) c_j(t) + m\mu c_{j+1}(t) \quad (j \geqq m).$$

Man kann auf Grund von (8.5.18) zeigen, daß das Gleichungssystem

$$-\lambda c_0 + \mu c_1 = 0, \tag{8.5.19}$$
$$\lambda c_{j-1} - (\lambda + j\mu) c_j + \mu(j+1) c_{j+1} = 0 \qquad (j = 1, \ldots, m-1),$$
$$\lambda c_{j-1} - (\lambda + m\mu) c_j + \mu m c_{j+1} = 0 \qquad (j = m, m+1, \ldots)$$

existiert, wobei $c_j = \lim\limits_{t\to\infty} c_j(t)$ ist. Als Lösungen dieses Gleichungssystems erhalten wir

$$c_j = \begin{cases} \dfrac{1}{j!}\left(\dfrac{\lambda}{\mu}\right)^j c_0 & (j = 1, \ldots, m), \\ \dfrac{1}{m!\, m^{j-m}}\left(\dfrac{\lambda}{\mu}\right)^j c_0 & (j \geqq m+1). \end{cases} \tag{8.5.20}$$

Es ist

$$\frac{1}{c_0}\sum_{j=0}^{\infty} c_j = \sum_{j=0}^{m}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \frac{m^m}{m!}\sum_{j=m+1}^{\infty}\left(\frac{\lambda}{m\mu}\right)^j. \tag{8.5.21}$$

Ist $m \leqq \dfrac{\lambda}{\mu}$, so strebt die rechte Seite der letzten Gleichung gegen ∞. In diesem Fall existiert keine Grenzverteilung; denn die Beziehung

$$\frac{1}{c_0}\sum_{j=0}^{\infty} c_j = \infty$$

kann nur gelten, wenn $c_0 \neq 0$ und $\sum\limits_{j=0}^{\infty} c_j = \infty$ ist (die c_j sind dann keine Wahrscheinlichkeiten) oder wenn $c_0 = 0$ ist. Dann folgt[1]) aber aus den Formeln (8.5.20) $c_j = 0$ $(j = 0, 1, 2, \ldots)$.

Im letzten Fall ist die Wahrscheinlichkeit dafür, daß die Anzahl der auf Entlastung wartenden Signale unendlich wird, gleich Eins.

Ist $\dfrac{\lambda}{\mu} < m$, so setzen wir $\sum\limits_{j=0}^{\infty} c_j = 1$ und erhalten aus (8.5.21)

$$c_0 = \frac{1}{\sum\limits_{k=0}^{m}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \frac{1}{m!}\frac{1}{m-\frac{\lambda}{\mu}}\left(\frac{\lambda}{\mu}\right)^{m+1}}$$

und schließlich

$$c_j = \begin{cases} \dfrac{\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j}{\sum\limits_{k=0}^{m}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \frac{1}{m!}\frac{1}{m-\frac{\lambda}{\mu}}\left(\frac{\lambda}{\mu}\right)^{m+1}} & (j = 0, 1, 2, \ldots, m), \\[2ex] \dfrac{\frac{1}{m!}\frac{1}{m^{j-m}}\left(\frac{\lambda}{\mu}\right)^j}{\sum\limits_{k=0}^{m}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \frac{1}{m!}\frac{1}{m-\frac{\lambda}{\mu}}\left(\frac{\lambda}{\mu}\right)^{m+1}} & (j \geqq m+1). \end{cases} \tag{8.5.22}$$

---

[1]) In diesem Fall gilt $\lim\limits_{t\to\infty} c_j(t) = 0$ $(j = 0, 1, 2, \ldots)$, obwohl $\lim\limits_{t\to\infty}\dfrac{1}{c_0(t)}\sum\limits_{j=0}^{\infty} c_j(t) = \infty$ ist.

Obwohl wir es in praktischen Fragen immer mit endlich vielen Signalen zu tun haben, können doch die praktischen Konsequenzen der Tatsache, daß $m \leqq \frac{\lambda}{\mu}$ ist, ernstlich die Tätigkeit einer „Zentrale" stören, da endlose Schlangen entstehen können. Die Anzahl der „Leitungen" darf also nicht zu klein sein.

Wir bezeichnen jetzt mit Z die Wartezeit eines Signals auf das Freiwerden einer Leitung. Dabei seien im Augenblick t des Aufleuchtens dieses Signals alle Leitungen belegt, d. h., es sei $Y_t \geqq m$. Unsere Aufgabe besteht nun darin, die Wahrscheinlichkeit $P(Z > z)$ für $z \geqq 0$ zu berechnen. Wir beschränken uns auf große Werte von t, für die man die Grenzformeln (8.5.22) benutzen kann. Da Y_t im Augenblick, wenn das betrachtete Signal aufleuchtet, nur gleich einer Zahl sein kann, die nicht kleiner als m ist, erhalten wir nach dem Satz über die totale Wahrscheinlichkeit

$$P(Z > z) = \sum_{j=m}^{\infty} c_j P(Z > z \mid Y_t = j). \tag{8.5.23}$$

Für $Y_t = j$ und $j \geqq m$ ist die Anzahl der wartenden Signale gleich $j - m$. Besteht die Ungleichung $Z > z$, so erlöschen in der Zeit z, gerechnet vom Aufleuchten des betrachteten Signals an, höchstens $j - m$ Signale. Aus der Voraussetzung, daß die Signaldauer eine Exponentialverteilung mit dem Parameter μ besitzen soll, und aus den übrigen Voraussetzungen über die erlöschenden Signale, die am Anfang dieses Paragraphen formuliert wurden, sowie schließlich aus der Tatsache, daß die Anzahl der belegten Leitungen stets gleich m ist (jede freiwerdende Leitung wird doch sofort vom nächsten wartenden Signal belegt), folgt, daß der Strom der erlöschenden Signale einen Poissonschen Prozeß mit der Intensität $m\mu$ bildet. Dann ist nach der Formel (8.3.4) die Wahrscheinlichkeit dafür, daß l Signale in der Zeit z erlöschen, gleich

$$\frac{(\mu m z)^l}{l!} \exp(-\mu m z);$$

die Wahrscheinlichkeit $P(Z > z \mid Y_t = j)$ dafür, daß in der Zeit z höchstens $j - m$ Signale erlöschen, ist also gleich

$$P(Z > z \mid Y_t = j) = \exp(-\mu m z) \sum_{l=0}^{j-m} \frac{(\mu m z)^l}{l!}. \tag{8.5.24}$$

Wir schreiben (8.5.22) für $j \geqq m + 1$ in der Form

$$c_j = c_m \frac{1}{m^{j-m}} \left(\frac{\lambda}{\mu}\right)^{j-m}. \tag{8.5.25}$$

Aus (8.5.23) bis (8.5.25) erhalten wir

$$
\begin{aligned}
P(Z > z) &= \sum_{j=m}^{\infty} c_j \sum_{l=0}^{j-m} \exp(-\mu m z) \frac{(\mu m z)^l}{l!} \\
&= \exp(-\mu m z) \sum_{j=m}^{\infty} c_m \frac{1}{m^{j-m}} \left(\frac{\lambda}{\mu}\right)^{j-m} \sum_{l=0}^{j-m} \frac{(\mu m z)^l}{l!} \\
&= c_m \exp(-\mu m z) \sum_{l=0}^{\infty} \frac{(\mu m z)^l}{l!} \sum_{j=m+l}^{\infty} \left(\frac{\lambda}{\mu m}\right)^{j-m} \\
&= c_m \exp(-\mu m z) \sum_{l=0}^{\infty} \frac{(\lambda z)^l}{l!} \sum_{j=m+l}^{\infty} \left(\frac{\lambda}{\mu m}\right)^{j-m-l} \\
&= \frac{c_m \exp(-\mu m z)}{1 - \dfrac{\lambda}{\mu m}} \sum_{l=0}^{\infty} \frac{(\lambda z)^l}{l!} = \frac{c_m}{1 - \dfrac{\lambda}{\mu m}} \exp[-(m\mu - \lambda) z].
\end{aligned}
$$

Als endgültiges Resultat haben wir also

$$
P(Z > z) = \frac{\dfrac{c_0}{m!} \left(\dfrac{\lambda}{\mu}\right)^m \exp[-(m\mu - \lambda) z]}{1 - \dfrac{\lambda}{\mu m}} \tag{8.5.26}
$$

erhalten.

D. Für das weitere Studium der Geburts- und Todesprozesse verweisen wir den Leser auf die Arbeiten von Karlin und McGregor [1] bis [3], D. G. Kendall [2] und Urbanik [2] sowie auf das Buch von Gnedenko, Beljajew und Solowjew [1]. Eine ausführliche Behandlung der Wirkungsweise der „Zentralen" im weitesten Sinne dieses Begriffs findet man in einer Arbeit von Chintschin [7] (siehe auch Karlin und McGregor [4]). Eine umfassende Übersicht über Ergebnisse der Warteschlangentheorie findet man in den Büchern von Riordan [1], Syski [1] und Takács [6].

Beim Geburts- und Todesprozeß haben wir die „Lebensdauer" eines Signals betrachtet. In praktischen Anwendungen ist es mitunter erforderlich, gewisse andere Merkmale eines Signals neben der „Lebensdauer" zu betrachten, z. B. dessen Amplitude, die eine Zufallsvariable sein kann. Mit anderen Worten, der ursprüngliche Signalprozeß erzeugt einen neuen stochastischen Prozeß, den man einen *sekundären Prozeß* nennt. Takács [1] bis [4] hat unter schwächeren Voraussetzungen über den ursprünglichen Signalprozeß und das Leben von Signalen viele wichtige Ergebnisse erzielt.

Es lassen sich auch etwas allgemeinere Modelle von Geburts- und Todesprozessen betrachten, z. B. solche, in denen jedes existierende Element, nennen wir es etwa

„Partikel", eine oder auch mehrere neue Partikeln erzeugen kann. Man kann auch Modelle betrachten, in denen Partikeln verschiedener Typen unterschieden werden und es darum geht, die Wahrscheinlichkeitsverteilung für die Anzahl der Teilchen eines speziellen Typs zum Zeitpunkt t zu ermitteln. Wir nennen solche Prozesse *Verzweigungsprozesse*. Dieser Problematik sind zahlreiche Arbeiten gewidmet, insbesondere die Arbeiten von KOLMOGOROFF und DMITRIEW [1], HARRIS [1], SEWASTJANOW [1] sowie WOODS und BHARUCHA-REID [1]. Man vergleiche hierzu auch das Buch von BHARUCHA-REID [2].

Ein Verzweigungsprozeß mit diskreter oder stetiger Zeit, in dem die Zufallsvariablen stetig sind, wurde von JIŘINA [1] untersucht. MOYAL [1], [2] untersuchte Verzweigungsprozesse mit Zufallsvariablen eines sehr allgemeinen Typs.

8.6. Der Pólyasche Prozeß

In 5.13 beschäftigten wir uns mit der zusammengesetzten Pólyaschen Verteilung und leiteten die Formel (5.13.8) her, die die Wahrscheinlichkeitsfunktion einer Zufallsvariablen mit negativer Binomialverteilung angibt. Wir gingen von einer Zufallsvariablen X mit einer durch (5.13.4) gegebenen Poissonschen Verteilung aus. Dabei hatte der darin vorkommende Parameter λ eine durch (5.13.5) gegebene Gammaverteilung. Jetzt ersetzen wir in der Herleitung der Formel (5.13.7) die Zufallsvariable X durch den Poissonschen Prozeß X_t, d. h., wir ersetzen (5.13.4) durch

$$P(X_t = k \mid \lambda) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t) \qquad (k = 0, 1, \ldots), \tag{8.6.1}$$

wobei $\lambda > 0, t > 0$ ist. Dann erhalten wir statt (5.13.7) für $k = 1, 2, \ldots$ die Formel

$$P_k(t) = P(X_t = k) = \frac{v(v+1)\cdots[v+(k-1)]}{k!} \left(\frac{t}{a+t}\right)^k P_0(t),$$

wobei

$$P_0(t) = P(X_t = 0) = \left(\frac{a}{a+t}\right)^v$$

ist. Wir setzen $a = v = \dfrac{1}{d}$ und erhalten

$$P_0(t) = (1 + td)^{-\frac{1}{d}},$$

$$P_k(t) = \frac{1(1+d)\cdots[1+(k-1)d]}{k!} \left(\frac{t}{1+td}\right)^k P_0(t). \quad (k=1,2,\ldots). \tag{8.6.2}$$

Die Funktionen $P_k(t)$ erfüllen die Anfangsbedingungen

$$P_0(0) = 1, \quad P_k(0) = 0 \qquad (k = 1, 2, 3, \ldots).$$

Differenzieren wir die Formeln (8.6.2), so erhalten wir für die Funktionen $P_k(t)$ das Differentialgleichungssystem

$$\begin{aligned} P_0'(t) &= -\frac{1}{1+td} P_0(t), \\ P_k'(t) &= -\frac{1+kd}{1+td} P_k(t) + \frac{1+(k-1)d}{1+td} P_{k-1}(t) \qquad (k = 1, 2, \ldots). \end{aligned} \tag{8.6.3}$$

Dieses Gleichungssystem hat dieselbe Form wie das System (8.4.4), wobei $i = 0$, $V_k(t) = P_k(t)$ und

$$\lambda_k(t) = \frac{1+kd}{1+td} \tag{8.6.4}$$

war. Also ist der betrachtete Prozeß ein Geburtsprozeß, den wir schon aus 8.4 kennen, mit dem Unterschied, daß $\lambda_k(t)$ hier von t abhängt. Wenn wir den Parameter t als Zeit betrachten, ist der behandelte Prozeß nicht mehr homogen (vgl. Beispiel 7.3.5). Wählt man aber den Parameter τ, der mit dem Parameter t durch die Beziehung

$$t = \frac{\exp(d\tau) - 1}{d}$$

verknüpft ist, als „Zeit", so wird der Prozeß homogen. Wir erhalten nämlich nach entsprechenden Umformungen aus (8.6.3)

$$\begin{aligned} \frac{dP_0(\tau)}{d\tau} &= -P_0(\tau), \\ \frac{dP_k(\tau)}{d\tau} &= -(1+kd)P_k(\tau) + [1+(k-1)d]P_{k-1}(\tau) \quad (k = 1, 2, \ldots). \end{aligned} \tag{8.6.5}$$

Die Koeffizienten $\lambda_k = 1 + kd$ sind hier unabhängig von τ; der Prozeß ist also homogen. Diese Zeitänderung darf nicht mißverstanden werden, die Deutung des Parameters als „Zeit" war ja nur eine Verabredung.

Man kann zeigen, daß die Funktionen (8.6.2) diejenigen Lösungen des Systems (8.6.3) sind, die den Anfangsbedingungen $P_0(0) = 1$, $P_k(0) = 0\,(k \geqq 1)$ genügen.

Den Mittelwert $E(X_t)$ und die Dispersion $D^2(X_t)$ finden wir aus den Formeln (5.13.10). Es ergibt sich

$$E(X_t) = \frac{1}{d}\left(\frac{td}{1+td} : \frac{1}{1+td}\right) = t, \tag{8.6.6}$$

$$D^2(X_t) = t(1+td). \tag{8.6.7}$$

Definition 8.6.1. Einen Prozeß $\{X_t, 0 \leqq t < \infty\}$, dessen Wahrscheinlichkeitsfunktionen $P(X_t = k)$ durch die Formeln (8.6.2) gegeben wird, nennt man einen *Pólyaschen Prozeß*.

Ein Pólyascher Prozeß ist ein Spezialfall der sogenannten *zusammengesetzten Poissonschen Prozesse*. Man erhält diese ebenso wie die Pólyaschen Prozesse: Das in der Formel (8.6.1) vorkommende λ ist eine Zufallsvariable mit der Verteilungsfunktion $U(\lambda)$. Ist also insbesondere $U(\lambda)$ durch die Formel (5.13.5) gegeben, so haben wir es mit einem Pólyaschen Prozeß zu tun. Ausführlich sind die zusammengesetzten Poissonschen Prozesse bei Lundberg [1] behandelt. Vorher wurden sie bereits von Chintschin [2] untersucht.

Der Name Pólyascher Prozeß stammt vom Pólyaschen Versuchsschema und von der in 5.4 besprochenen Pólyaschen Verteilung. So bewies Pólya [1], daß die rechte Seite von (5.4.2) für $n \to \infty$ gegen die durch die Formeln (8.6.2) definierte Funktion $P_k(t)$ konvergiert, wenn $\lim_{n\to\infty} np = t$ und $\lim_{n\to\infty} n\alpha = td$ ist.[1]) Den Pólyaschen Prozeß kann man also auf zwei verschiedene Arten herleiten.

Arley [1] wandte den Pólyaschen Prozeß in der Theorie der kosmischen Strahlung an. Lundberg [1] untersuchte viele Anwendungen des Pólyaschen Prozesses in der Unfall- und Krankenversicherungsstatistik. Das folgende Beispiel haben wir dem zitierten Lundbergschen Buch entnommen.

Beispiel 8.6.1. Die Versicherungsgesellschaft Sinkförsäkrings-A.B. Eir verpflichtet sich, den Versicherten im Fall der vollständigen Arbeitsunfähigkeit infolge Krankheit oder Unfall eine Unterstützung auszuzahlen. Im Lundbergschen Beobachtungsmaterial hatte (mit wenigen Ausnahmen) jeder Versicherte nur eine einzige Versicherungspolice.

Die statistischen Angaben, die die Anzahl dieser Policen betreffen, können also als Angaben über die Versicherten angesehen werden. Die Gesamtzahl der Policen betrug 1417. Wir betrachten die Anzahl k der Krankheits- bzw. Unfallmeldungen in einem Zeitraum von 10 Jahren, nämlich vom 3. bis 12. Versicherungsjahr. Die Angaben sind in Tabelle 8.6.1 aufgeführt. Dabei ist n_k die Anzahl der Versicherten, die im betrachteten Zeitabschnitt k-mal einen Unfall oder Krankheitsfall meldeten; P_k ist die nach der Formel (8.6.2) berechnete Wahrscheinlichkeit. Hierbei wurden die Parameter t und d aus dem Beobachtungsmaterial nach den Formeln (8.6.6) und (8.6.7) berechnet, und zwar $t = 2{,}9273$ und $d = 0{,}9416$.

Wie man aus Tabelle 8.6.1 ersehen kann, stimmen die theoretischen Wahrscheinlichkeiten gut mit den beobachteten Häufigkeiten überein.

[1]) Wir gebrauchen hier die Bezeichnungen aus 5.4.

Tabelle 8.6.1.

k	$\frac{n_k}{1417}$	P_k
0	0,2286	0,2453
1	0,2110	0,1911
2	0,1475	0,1446
3	0,1066	0,1083
4	0,0854	0,0807
5	0,0480	0,0599
6	0,0416	0,0445
7	0,0346	0,0329
8	0,0310	0,0243
9	0,0191	0,0180
10	0,0085	0,0133
11	0,0106	0,0098
12	0,0078	0,0072
13	0,0056	0,0053
14	0,0028	0,0039
15	0,0028	0,0029
16	0,0085	0,0080
	1,0000	1,0000

8.7. Die Kolmogoroffschen Gleichungen

In 8.3 bis 8.6 betrachteten wir diskrete homogene Prozesse, bei denen sich ein Zustand nur durch unmittelbaren Übergang zum Nachbarzustand ändern konnte. In diesem Paragraphen wollen wir diskrete Markoffsche Prozesse betrachten, bei denen jeder Zustand in einen beliebigen anderen übergehen kann.

Es sei $\{X_t, 0 \leqq t < \infty\}$ ein Markoffscher Prozeß mit höchstens abzählbar vielen Zuständen $i = 0, 1, 2, \ldots$, und $p_{ij}(\tau, t)$ $(0 \leqq \tau < t < \infty)$ sei die Übergangswahrscheinlichkeitsfunktion dieses Prozesses. Diese Funktion erfüllt dann die Beziehungen (8.2.6) bis (8.2.8). Die Chapman-Kolmogoroffsche Gleichung schreiben wir hier in der Form

$$p_{ij}(\tau, t) = \sum_k p_{ik}(\tau, s)\, p_{kj}(s, t); \tag{8.7.1}$$

dabei ist $\tau < s < t$. Wir wollen ein Differentialgleichungssystem für die Funktionen $p_{ij}(\tau, t)$ suchen. Dazu führen wir die Bezeichnungen

$$q_i(t) = \lim_{\Delta t \to 0} \frac{1 - p_{ii}(t, t + \Delta t)}{\Delta t} \qquad (i = 0, 1, \ldots), \tag{8.7.2}$$

$$q_{ij}(t) = \lim_{\Delta t \to 0} \frac{p_{ij}(t, t + \Delta t)}{\Delta t} \qquad (i, j = 0, 1, \ldots;\ i \neq j) \tag{8.7.3}$$

ein und nehmen an, daß in beiden Beziehungen die Konvergenz gleichmäßig in t ist.

Die Größe Δt kann in den Formeln (8.7.2) bis (8.7.3) positiv oder negativ sein. Diese Beziehungen sind den folgenden Gleichungen, die für kleine Δt gelten, äquivalent:

$$p_{ii}(t, t + \Delta t) = 1 - q_i(t)\,\Delta t + o(\Delta t), \tag{8.7.2'}$$

$$p_{ij}(t, t + \Delta t) = q_{ij}(t)\,\Delta t + o(\Delta t). \tag{8.7.3'}$$

Die Funktionen $q_i(t)$ und $q_{ij}(t)$ heißen *Intensitätsfunktionen*.

Den folgenden Satz 8.7.1 bewies KOLMOGOROFF in seiner Arbeit [3], die bei der Entwicklung der Theorie der Markoffschen Prozesse eine grundlegende Rolle spielte. In dem Satz werden zwei nach KOLMOGOROFF benannte Differentialgleichungssysteme hergeleitet, denen die Funktionen $p_{ij}(\tau, t)$ genügen. Das eine System (8.7.4) erhält man, wenn man j und t als fest und i und τ als veränderlich ansieht, das zweite (8.7.6), wenn man i und τ als fest, dagegen aber j und t als veränderlich ansieht. Da $\tau < t$ ist, heißen die ersten Gleichungen (8.7.4) *retrospektiv*, während die zweiten (8.7.6) *prospektiv* genannt werden.

Satz 8.7.1. *Es seien $p_{ij}(\tau, t)$ die Übergangswahrscheinlichkeitsfunktionen eines Markoffschen Prozesses mit höchstens abzählbar vielen Zuständen $i = 0, 1, 2, \ldots$ Die Intensitätsfunktionen $q_i(t)$ und $q_{ij}(t)$ mögen existieren und stetig sein. Dann gilt:*

1. *Die Funktionen $p_{ij}(\tau, t)$ erfüllen das Differentialgleichungssystem*

$$\frac{\partial p_{ij}(\tau, t)}{\partial \tau} = q_i(\tau)\, p_{ij}(\tau, t) - \sum_{k \neq i} q_{ik}(\tau)\, p_{kj}(\tau, t) \tag{8.7.4}$$

bei den Anfangsbedingungen

$$p_{ij}(t, t) = \begin{cases} 1 & \text{für } j = i, \\ 0 & \text{für } j \neq i. \end{cases} \tag{8.7.5}$$

2. *Wenn wir außerdem noch voraussetzen, daß die Konvergenz in (8.7.3) für jedes feste j gleichmäßig bezüglich i ist, so erfüllen die Funktionen $p_{ij}(\tau, t)$ das Differentialgleichungssystem*

$$\frac{\partial p_{ij}(\tau, t)}{\partial t} = -p_{ij}(\tau, t)\, q_j(t) + \sum_{k \neq j} p_{ik}(\tau, t)\, q_{kj}(t), \tag{8.7.6}$$

unter den Anfangsbedingungen

$$p_{ij}(\tau, \tau) = \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases} \tag{8.7.7}$$

Beweis. Wir beweisen die Behauptung 2. Dazu schreiben wir (8.7.1) in der Form

$$p_{ij}(\tau, t + \Delta t) = \sum_k p_{ik}(\tau, t)\, p_{kj}(t, t + \Delta t).$$

Dann erhalten wir

$$\frac{p_{ij}(\tau, t + \Delta t) - p_{ij}(\tau, t)}{\Delta t} = \frac{1}{\Delta t}\left[\sum_k p_{ik}(\tau, t)\, p_{kj}(t, t + \Delta t) - p_{ij}(\tau, t)\right]$$

$$= \sum_{k \neq j} p_{ik}(\tau, t) \frac{p_{kj}(t, t + \Delta t)}{\Delta t} - p_{ij}(\tau, t) \frac{1 - p_{jj}(t, t + \Delta t)}{\Delta t}.$$

Aus den Formeln (8.7.2) bis (8.7.3), der absoluten Konvergenz der Reihe $\sum_k p_{ik}(\tau, t)$ und den Voraussetzungen über die gleichmäßige Konvergenz von (8.7.3) bezüglich i (hier bezüglich k) bei festem j folgt dann

$$\frac{\partial p_{ij}(\tau, t)}{\partial t} = \lim_{\Delta t \to 0} \frac{p_{ij}(\tau, t + \Delta t) - p_{ij}(\tau, t)}{\Delta t}$$

$$= \sum_{k \neq j} p_{ik}(\tau, t) \lim_{\Delta t \to 0} \frac{p_{kj}(t, t + \Delta t)}{\Delta t} - p_{ij}(\tau, t) \lim_{\Delta t \to 0} \frac{1 - p_{jj}(t, t + \Delta t)}{\Delta t}$$

$$= -p_{ij}(\tau, t)\, q_j(t) + \sum_{k \neq j} p_{ik}(\tau, t)\, q_{kj}(t).$$

Damit haben wir die Behauptung 2 bewiesen. Der Beweis der Behauptung 1 ist analog.

Man kann zeigen, daß die Gleichungssysteme (8.7.4) und (8.7.6) identische Lösungen $p_{ij}(\tau, t)$ besitzen, die die Formeln (8.2.6) bis (8.2.8) und die vorausgesetzte Anfangsbedingung erfüllen, und daß die diesen Lösungen entsprechenden Intensitätsfunktionen existieren und stetig sind. Den Beweis findet der Leser in der Arbeit von Feller [1] und für den Spezialfall eines homogenen Prozesses in der schon erwähnten Arbeit von Kolmogoroff [3]. Wenn der Prozeß nur endlich viele Zustände hat, ist offenbar die Voraussetzung über die gleichmäßige Konvergenz in (8.7.3) bezüglich i bei festem j überflüssig.

Falls der Prozeß homogen ist, sind die Intensitätsfunktionen, die durch die Formeln (8.7.2) bis (8.7.3) gegeben werden, von t unabhängig; man kann also einfach q_i und q_{ij} schreiben und diese Größen *Intensitäten* nennen. Die Gleichungen (8.7.4) und (8.7.6) haben dann die Form

$$\frac{d p_{ij}(t)}{dt} = q_i p_{ij}(t) - \sum_{k \neq i} q_{ik} p_{kj}(t), \tag{8.7.4'}$$

$$\frac{d p_{ij}(t)}{dt} = -p_{ij}(t) q_j + \sum_{k \neq j} p_{ik}(t)\, q_{kj} \tag{8.7.6'}$$

mit den Anfangsbedingungen

$$p_{ij}(0) = \begin{cases} 1 & \text{für} \quad j = i, \\ 0 & \text{für} \quad j \neq i. \end{cases} \tag{8.7.5'}$$

Wir weisen darauf hin, daß alle Differentialgleichungssysteme, die wir in 8.3 bis 8.5 erhalten haben, Spezialfälle des Systems (8.7.6′) sind.

In Satz 8.7.1 nahmen wir an, daß die Beziehungen (8.7.2) und (8.7.3) erfüllt sind, daß also die Intensitätsfunktionen $q_i(t)$ und $q_{ij}(t)$ existieren und endlich sind. Für homogene Markoffsche Prozesse folgt dies aber, wie Doob [2] und Kolmogoroff [15] gezeigt haben, aus der Voraussetzung der Stetigkeit der Funktion $p_{ij}(t)$ im Punkt $t = 0$, d. h. aus der Voraussetzung

$$\lim_{t \to 0} p_{ii}(t) = 1 \qquad (i = 0, 1, 2, \ldots). \tag{8.7.8}$$

Setzt man noch zusätzlich voraus, daß die Konvergenz in der Beziehung (8.7.8) gleichmäßig in i ist, so sind die Intensitäten q_i und q_{ij} endlich. Hat aber der betrachtete Prozeß nur endlich viele Zustände, so folgt die Existenz endlicher Intensitäten aus der Beziehung (8.7.8). Die Voraussetzung (8.7.8) ist wegen der Anfangsbedingungen (8.7.5′) ganz selbstverständlich.

Der Existenz der Intensitätsfunktionen $q_i(t)$ und $q_{ij}(t)$ sowie deren Eigenschaften sind auch Arbeiten von Austin [1], [2] und Juschkewitsch [1] gewidmet.

8.8. Rein unstetige und rein stetige Prozesse

A. In 8.3 bis 8.7 untersuchten wir Prozesse mit höchstens abzählbar vielen Zuständen. Nun nehmen wir an, daß die Zufallsvariablen X_t eines Markoffschen Prozesses $\{X_t, 0 \leqq t < \infty\}$ beliebige reelle Werte annehmen können. Wir erinnern daran, daß für einen Markoffschen Prozeß und beliebige t_1, t_2 $(0 \leqq t_1 < t_2 < \infty)$ die bedingte Verteilungsfunktion $F(t_1, x, t_2, y) = P(X_{t_1} < y \mid X_{t_2} = x)$ gegeben ist. Dabei ist im Einklang mit unserer Voraussetzung die bedingte Wahrscheinlichkeit auf der rechten Seite der letzten Gleichung für jeden Wert von x bestimmt. Die Funktion $F(t_1, x, t_2, y)$ ist als Verteilungsfunktion wenigstens linksseitig stetig in y und genügt den Gleichungen

$$\lim_{y \to -\infty} F(t_1, x, t_2, y) = 0, \qquad \lim_{y \to \infty} F(t_1, x, t_2, y) = 1.$$

Außerdem sei $F(t_1, x, t_2, y)$ stetig bezüglich t_1 und t_2 und eine Bairesche Funktion (vgl. 2.4.A) in x. Die Chapman-Kolmogoroffsche Gleichung hat hier die Form

$$F(t_1, x, t_2, y) = \int_{-\infty}^{\infty} F(s, z, t_2, y)\, d_z F(t_1, x, s, z), \tag{8.8.1}$$

wobei $t_1 < s < t_2$ ist. Das Integral in (8.8.1) ist das Stieltjessche Integral von $F(s, z, t_2, y)$ nach der Verteilungsfunktion $F(t_1, x, s, z)$. Hat $F(t_1, x, t_2, y)$ bezüglich y eine Dichtefunktion $f(t_1, x, t_2, y)$, so lautet die Chapman-Kolmogoroffsche Gleichung

$$F(t_1, x, t_2, y) = \int_{-\infty}^{\infty} F(s, z, t_2, y)\, f(t_1, x, s, z)\, dz, \tag{8.8.1'}$$

wobei $t_1 < s < t_2$ ist. Die Gleichung (8.8.1) oder (8.8.1') kann man so begründen: Den Übergang vom Zustand x, in dem sich der Prozeß im Moment t_1 befindet, durch einen Zustand z im Augenblick s zu einem Zustand, der kleiner als y ist und in dem der Prozeß sich im Augenblick t_2 befindet, kann man in zwei Übergänge zerlegen: 1. Im Augenblick s befindet sich der Prozeß in einem Zustand, der zum „Intervall" $[z, z + dz)$ gehört; die Wahrscheinlichkeit dafür ist gleich $f(t_1, x, s, z)\, dz$. 2. Aus dem Zustand z, in dem sich der Prozeß im Moment s befindet, geht der Prozeß in einen Zustand über, der kleiner als y ist und den der Prozeß im Moment t_2 annimmt; die Wahrscheinlichkeit dieses Ereignisses ist gleich $F(s, z, t_2, y)$.

Wenn wir den verallgemeinerten Satz über die totale Wahrscheinlichkeit [vgl. Formel (2.7.7)] anwenden, erhalten wir (8.8.1'). Aus (8.8.1') folgt unmittelbar

$$f(t_1, x, t_2, y) = \int_{-\infty}^{\infty} f(s, z, t_2, y)\, f(t_1, x, s, z)\, dz. \tag{8.8.2}$$

Aus der Definition der Funktion $F(t_1, x, t_2, y)$ und der Voraussetzung, daß sie in t_1 und t_2 stetig sein soll, folgt

$$\lim_{t_1 \to t_2} F(t_1, x, t_2, y) = \begin{cases} 0 & \text{für} \quad y \leqq x, \\ 1 & \text{für} \quad y > x, \end{cases} \tag{8.8.3}$$

$$\lim_{t_2 \to t_1} F(t_1, x, t_2, y) = \begin{cases} 0 & \text{für} \quad y \leqq x, \\ 1 & \text{für} \quad y > x. \end{cases} \tag{8.8.4}$$

Dabei ist die Konvergenz in (8.8.3) linksseitig und in (8.8.4) rechtsseitig.

B. Wir führen jetzt den Begriff eines rein unstetigen Prozesses ein. Es sei $\{X_t, t \in I\}$ ein Markoffscher Prozeß, wobei wir nicht ausschließen, daß alle reellen Zahlen mögliche Zustände dieses Prozesses sein können. Es seien zwei Funktionen $q(t, x)$ und $P(t, x, y)$ durch

$$q(t, x) = \lim_{\Delta t \to 0} \frac{P(X_{t+\Delta t} - X_t \neq 0 \mid X_t = x)}{\Delta t},$$

$$P(t, x, y) = \lim_{\Delta t \to 0} P(X_{t+\Delta t} < y \mid X_t = x, \quad X_{t+\Delta t} - X_t \neq 0)$$

definiert. Die Funktion $q(t, x)$ ist nichtnegativ, und $P(t, x, y)$ ist als Funktion von y eine Verteilungsfunktion. Ein Markoffscher Prozeß heißt rein unstetig,

wenn er nur sprunghaft Änderungen erfahren kann, und zwar auf folgende Weise: Wenn im Augenblick t der Prozeß sich im Zustand x befindet, dann ist die Wahrscheinlichkeit dafür, daß der Prozeß im Augenblick $t + \Delta t$ in einem von x verschiedenen Zustand ist, gleich $q(t, x)\,\Delta t + o(\Delta t)$. Einen rein unstetigen Markoffschen Prozeß kann man analytisch folgendermaßen definieren:

Definition 8.8.1. Ein Markoffscher Prozeß $\{X_t, t \in I\}$ heißt *rein unstetig*, wenn identisch in t gilt:

$$F(t, x, t + \Delta t, y) = \begin{cases} 1 - q(t, x)\,\Delta t + q(t, x)\,P(t, x, y)\,\Delta t + o(\Delta t) & \text{für } x < y, \\ q(t, x)\,P(t, x, y)\,\Delta t + o(\Delta t) & \text{für } x \geqq y. \end{cases} \tag{8.8.5}$$

Die Deutung dieser Gleichung ist einfach: Ist $x < y$, so kann der Prozeß im Augenblick $t + \Delta t$ in einem Zustand kleiner als y sein, wenn er entweder im Augenblick t im Zustand x war und während des Zeitabschnitts $[t, t + \Delta t)$ keine Zustandsänderung erfolgt ist, oder wenn eine Zustandsänderung erfolgt, und zwar so, daß sich der Prozeß im Augenblick $t + \Delta t$ in einem Zustand kleiner als y befindet. Ist $x \geqq y$, so kann der Prozeß im Zeitpunkt $t + \Delta t$ nur dann in einem Zustand kleiner als y sein, wenn eine entsprechende Zustandsänderung eintritt.

Setzen wir $E(x, y) = 1$ für $x < y$, $E(x, y,) = 0$ für $x \geqq y$,
so können wir die Gleichung (8.8.5) in der folgenden Form schreiben:

$$\begin{aligned} F(t, x, t + \Delta t, y) = {} & [1 - q(t, x)\,\Delta t]\,E(x, y) \\ & + q(t, y)\,P(t, x, y)\,\Delta t + o(\Delta t). \end{aligned} \tag{8.8.5'}$$

Satz 8.8.1. *Es sei $\{X_t, 0 \leqq t < \infty\}$ ein rein unstetiger Markoffscher Prozeß, und es sei $F(t_1, x, t_2, y)$ eine bezüglich t_1 und t_2 stetige Bairesche Funktion in x. Ferner seien $P(t, x, y)$ und $q(t, x)$ bezüglich t stetige Bairesche Funktionen in x. Dann gelten die folgenden Behauptungen:*

1. *Die Verteilungsfunktion $F(t_1, x, t_2, y)$ genügt den beiden Integrodifferentialgleichungen*

$$\begin{aligned} \frac{\partial F(t_1, x, t_2, y)}{\partial t_1} = {} & q(t_1, x)\,[F(t_1, x, t_2, y) \\ & - \int_{-\infty}^{\infty} F(t_1, x, t_2, y)\,d_z P(t_1, x, z)], \end{aligned} \tag{8.8.6}$$

$$\begin{aligned} \frac{\partial F(t_1, x, t_2, y)}{\partial t_2} = {} & - \int_{-\infty}^{y} q(t_2, z)\,d_z F(t_1, x, t_2, z) \\ & + \int_{-\infty}^{\infty} q(t_2, z)\,P(t_2, z, y)\,d_z F(t_1, x, t_2, z); \end{aligned} \tag{8.8.6'}$$

2. *Die Gleichungen* (8.8.6) *und* (8.8.6′) *haben unter Voraussetzung der Anfangsbedingungen* (8.8.3) *und* (8.8.4) *nur eine Lösung* $F(t_1, x, t_2, y)$, *welche beiden Gleichungen gemeinsam ist und alle Eigenschaften besitzt, die in den Voraussetzungen dieses Satzes festgelegt wurden.*

Wir begnügen uns hier damit, die Beziehung (8.8.6′) zu beweisen, da der Beweis für die Beziehung (8.8.6) völlig analog verläuft. Auf den Beweis der zweiten Behauptung verzichten wir überhaupt. Der Leser findet ihn in der Arbeit von FELLER [1], von dem auch die Formulierung dieser Behauptung stammt.

Beweis. Unter Anwendung der Formeln (8.8.1) und (8.8.5′) sowie unter Berücksichtigung des Wertes von $E(z, y)$ erhalten wir für $\Delta t_2 > 0$

$$\begin{aligned} F(t_1, x, t_2 + \Delta t_2, y) &= \int_{-\infty}^{\infty} F(t, z, t_2 + \Delta t_2, y)\, d_z F(t_1, x, t_2, z) \\ &= \int_{-\infty}^{\infty} \{[1 - q(t_2, z)\,\Delta t_2]\, E(z, y) \\ &\quad + [q(t_2, z)\, P(t_2, z, y)\,\Delta t_2] + o(\Delta t_2)\}\, d_z F(t_1, x, t_2, z) \\ &= F(t_1, x, t_2, y) - \Delta t_2 \int_{-\infty}^{y} q(t_2, z)\, d_z F(t_1, x, t_2, z) \\ &\quad + \Delta t_2 \int_{-\infty}^{\infty} q(t_2, z)\, P(t_2, z, y)\, d_z F(t_1, x, t_2, z) + o(\Delta t_2). \end{aligned}$$

Es sei darauf hingewiesen, daß die Existenz sämtlicher hier auftretenden Integrale aus den über die Funktionen q und P getroffenen Voraussetzungen folgt. Aus den übrigen Gleichungen erhalten wir

$$\begin{aligned} &\lim_{\Delta t_2 \to 0} \frac{F(t_1, x, t_2 + \Delta t_2, y) - F(t_1, x, t_2, y)}{\Delta t_2} \\ &= -\int_{-\infty}^{y} q(t_2, z)\, d_z F(t_1, x, t_2, z) + \int_{-\infty}^{\infty} q(t_2, z)\, P(t_2, z, y)\, d_z F(t_1, x, t_2, z). \end{aligned}$$

Demnach existiert die rechtsseitige Ableitung.

Wiederholen wir diese Überlegung für $\Delta t_2 < 0$, so erhalten wir die linksseitige Ableitung, die sich als der rechtsseitigen gleich erweist. Daraus folgt die Existenz der Ableitung $\frac{\partial}{\partial t_2} F(t_1, x, t_2, y)$. Ferner sieht man, daß diese Ableitung durch (8.8.6′) gegeben ist.

Die Gleichung (8.8.6) heißt *retrospektive Gleichung*; sie wurde von FELLER [1] hergeleitet. Die Gleichung (8.8.6′) wurde (unter etwas anderen Voraussetzungen) von KOLMOGOROFF [3] hergeleitet und heißt *prospektive Gleichung*.

C. Wir kommen nun zur Behandlung eines rein stetigen Prozesses.

Definition 8.8.2. Ein Markoffscher Prozeß $\{X_t, t \in I\}$ heißt *rein stetig*, wenn für jedes $t \in I$ und jedes $\varepsilon > 0$ die folgenden Beziehungen gelten:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|y-x| \geqq \varepsilon} (y - x)^2 \, d_y F(t, x, t + \Delta t, y) = 0, \tag{8.8.7}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|y-x| < \varepsilon} (y - x) \, d_y F(t, x, t + \Delta t, y) = m(t, x), \tag{8.8.8}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|y-x| < \varepsilon} (y - x)^2 \, d_y F(t, x, t + \Delta t, y) = \sigma^2(t, x) > 0. \tag{8.8.9}$$

Dabei soll die Konvergenz in den letzten beiden Beziehungen gleichmäßig in x sein. Die Größe Δt kann beliebiges Vorzeichen haben.

Existiert die Dichtefunktion $f(t_1, x, t_2, y)$, so nehmen die oben angegebenen Formeln die folgende Gestalt an:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|y-x| \geqq \varepsilon} (y - x)^2 f(t, x, t + \Delta t, y) \, dy = 0, \tag{8.8.7'}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|y-x| < \varepsilon} (y - x) f(t, x, t + \Delta t, y) \, dy = m(t, x), \tag{8.8.8'}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|y-x| < \varepsilon} (y - x)^2 f(t, x, t + \Delta t, y) \, dy = \sigma^2(t, x) > 0. \tag{8.8.9'}$$

Wir bemerken, daß sich aus (8.8.7)

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|y-x| \geqq \varepsilon} d_y F(t, x, t + \Delta t, y) = 0$$

ergibt, während aus (8.8.7) und (8.8.9)

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (y - x)^2 \, d_y F(t, x, t + \Delta t, y) = \sigma^2(t, x) > 0$$

folgt. Die letzten beiden Formeln schreiben wir in der Gestalt

$$P(|X_{t+\Delta t} - X_t| \geqq \varepsilon \,|\, X_t = x) = o(\Delta t) \tag{8.8.10}$$

bzw.

$$E[(X_{t+\Delta t} - X_t)^2 | X_t = x] = \sigma^2(t, x)\,\Delta t + o(\Delta t). \tag{8.8.11}$$

Nach (8.8.10) ist die Wahrscheinlichkeit einer wesentlichen Zustandsänderung während eines kleinen Zeitintervalls sehr klein, und zwar von kleinerer Ordnung als Δt, und nach (8.8.11) ist, wie klein auch Δt ist, die Wahrscheinlichkeit dafür, daß $X_{t+\Delta t} - X_t \neq 0$ ist, positiv. Hier ist es umgekehrt wie bei einem rein unstetigen Prozeß. Dort war die Wahrscheinlichkeit, daß in einem kleinen Zeitabschnitt eine Zustandsänderung auftritt, klein; erfolgt aber eine Änderung, so kann sie groß sein.

D. Die in 8.3 bis 8.6 betrachteten Prozesse waren rein unstetig. Die Realisierungen dieser Prozesse wurden durch Treppenfunktionen dargestellt; es gibt aber auch rein unstetige Prozesse, deren Realisierungen bedeutend kompliziertere Funktionen sein können. Die Bezeichnung „rein stetige Prozesse" legt die Annahme nahe, daß die Realisierungen dieser Prozesse immer stetige Funktionen sind. Das ist in den meisten Fällen so, aber nicht immer. In 8.12 werden wir dieser Frage einige Bemerkungen widmen.

8.9. Der Wienersche Prozeß

Wir wollen uns jetzt mit einem rein stetigen Prozeß befassen, der große theoretische Bedeutung hat und viele physikalische Anwendungen besitzt.

Es sei $\{X_t, 0 \leqq t < \infty\}$ ein Markoffscher Prozeß, dessen mögliche Zustände alle reellen Zahlen sein sollen. Wir setzen voraus, daß der Prozeß rein stetig ist und daß für beliebige τ und t $(0 \leqq \tau < t < \infty)$ die bedingte Verteilungsfunktion $F(\tau, x, t, y)$ die Form $F(\tau, t, u)$ mit $u = y - x$ hat. $F(\tau, t, u)$ ist also die Wahrscheinlichkeit dafür, daß der Zuwachs in der Zeit von τ bis t kleiner als u ist. Die Funktion $F(\tau, t, u)$ ist nach Voraussetzung stetig in t und τ und hat die Dichtefunktion $f(\tau, t, u)$. Die Beziehungen (8.8.2) bis (8.8.4) haben hier die Form

$$f(\tau, t, u) = \int_{-\infty}^{\infty} f(\tau, s, u - w)\, f(s, t, w)\, dw, \tag{8.9.1}$$

wobei $\tau < s < t$ ist,

$$\lim_{\tau \to t} F(\tau, t, u) = \begin{cases} 0 & \text{für} \quad u \leqq 0, \\ 1 & \text{für} \quad u > 0; \end{cases} \tag{8.9.2}$$

$$\lim_{t \to \tau} F(\tau, t, u) = \begin{cases} 0 & \text{für} \quad u \leqq 0, \\ 1 & \text{für} \quad u > 0, \end{cases} \tag{8.9.3}$$

während die Beziehungen (8.8.7′) bis (8.8.9′) die Gestalt

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t} \int\limits_{|u|\geqq\varepsilon} u^2 f(t, t+\Delta t, u)\, du = 0, \tag{8.9.4}$$

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t} \int\limits_{|u|<\varepsilon} u f(t, t+\Delta t, u)\, du = m(t), \tag{8.9.5}$$

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t} \int\limits_{|u|<\varepsilon} u^2 f(t, t+\Delta t, u)\, du = \sigma^2(t) > 0 \tag{8.9.6}$$

haben; dabei ist $\varepsilon > 0$ beliebig.

Wir nehmen außerdem an:

(A) Für alle τ, t und u existieren die partiellen Ableitungen

$$\frac{\partial f(\tau, t, u)}{\partial u}, \qquad \frac{\partial^2 f(\tau, t, u)}{\partial u^2}$$

und sind stetig in allen Argumenten; außerdem ist für jedes feste t (oder τ) die Ableitung $\dfrac{\partial^2 f}{\partial u^2}$ gleichmäßig stetig und gleichmäßig beschränkt bezüglich u.

Satz 8.9.1. *Es sei* $\{X_t,\ 0 \leqq t < \infty\}$ *ein Markoffscher Prozeß, für den die Formeln* (8.9.1) *bis* (8.9.6) *sowie die Voraussetzung* (A) *gelten. Die Dichtefunktion* $f(\tau, t, u)$ *erfüllt dann die folgende partielle Differentialgleichung zweiter Ordnung vom parabolischen Typus:*

$$\frac{\partial f}{\partial t} = -m(t)\, \frac{\partial f}{\partial u} + \frac{1}{2}\, \sigma^2(t)\, \frac{\partial^2 f}{\partial u^2}. \tag{8.9.7}$$

Beweis. Wir schreiben (8.9.1) in der Form

$$f(\tau, t+\Delta t, u) = \int\limits_{-\infty}^{\infty} f(\tau, t, u-w)\, f(t, t+\Delta t, w)\, dw \tag{8.9.8}$$

und entwickeln $f(\tau, t, u-w)$ in eine Taylorsche Reihe:

$$\begin{aligned} f(\tau, t, u-w) = f(\tau, t, u) - w\, \frac{\partial f}{\partial u} + \frac{1}{2}\, w^2\, \frac{\partial^2 f}{\partial u^2} \\ + \frac{1}{2}\, w^2 \left\{ \left[\frac{\partial^2 f}{\partial u^2}\right]_{u-\vartheta w} - \frac{\partial^2 f}{\partial u^2} \right\} \qquad (0 < \vartheta < 1). \end{aligned} \tag{8.9.9}$$

Diese Entwicklungen setzen wir in (8.9.8) ein und erhalten

$$\frac{f(\tau, t+\Delta t, u)}{\Delta t} = \frac{f(\tau, t, u)}{\Delta t} \int_{-\infty}^{\infty} f(t, t+\Delta t, w)\, dw$$
$$- \frac{\partial f}{\partial u} \frac{1}{\Delta t} \int_{-\infty}^{\infty} w f(t, t+\Delta t, w)\, dw \tag{8.9.10}$$
$$+ \frac{1}{2} \frac{\partial^2 f}{\partial u^2} \frac{1}{\Delta t} \int_{-\infty}^{\infty} w^2 f(t, t+\Delta t, w)\, dw + \frac{1}{2\Delta t} J,$$

wobei

$$J = \int_{-\infty}^{\infty} w^2 \left\{ \left[\frac{\partial^2 f}{\partial u^2}\right]_{u-\vartheta w} - \frac{\partial^2 f}{\partial u^2} \right\} f(t, t+\Delta t, w)\, dw \tag{8.9.11}$$

ist. Wir bemerken, daß

$$\int_{-\infty}^{\infty} f(t, t+\Delta t, w)\, dw = 1 \tag{8.9.12}$$

ist. Ferner folgen aus den Formeln (8.9.4) bis (8.9.6) die Gleichungen

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} w f(t, t+\Delta t, w)\, dw$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|w|<\varepsilon} w f(t, t+\Delta t, w)\, dw = m(t), \tag{8.9.13}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} w^2 f(t, t+\Delta t, w)\, dw$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|w|<\varepsilon} w^2 f(t, t+\Delta t, w)\, dw = \sigma^2(t). \tag{8.9.14}$$

Wir spalten das Integral J in zwei Integrale J_1 und J_2 auf:

$$J_1 = \int_{|w|<\varepsilon} w^2 \left\{ \left[\frac{\partial^2 f}{\partial u^2}\right]_{u-\vartheta w} - \frac{\partial^2 f}{\partial u^2} \right\} f(t, t+\Delta t, w)\, dw, \tag{8.9.15}$$

$$J_2 = \int_{|w|\geqq\varepsilon} w^2 \left\{ \left[\frac{\partial^2 f}{\partial u^2}\right]_{u-\vartheta w} - \frac{\partial^2 f}{\partial u^2} \right\} f(t, t+\Delta t, w)\, dw. \tag{8.9.16}$$

Auf Grund der Voraussetzung über die gleichmäßige Stetigkeit der Ableitung $\frac{\partial^2 f}{\partial u^2}$ in u kann der Ausdruck in der geschweiften Klammer von (8.9.15) für jedes Δt beliebig klein gemacht werden, wenn ε genügend klein ist. Daraus folgt $J_1 = o(\Delta t)$. Wegen der gleichmäßigen Beschränktheit der Ableitung $\frac{\partial^2 f}{\partial u^2}$ in u und auf Grund der Formel (8.9.4) erhalten wir, daß auch $J_2 = o(\Delta t)$ gilt. Also ist

$$J = o(\Delta t). \tag{8.9.17}$$

Setzen wir in (8.9.10) die Formeln (8.9.12) bis (8.9.14) und (8.9.17) ein, so sehen wir, daß die rechtsseitige Ableitung $\frac{\partial f}{\partial t}$ existiert; der analoge Schluß gilt für $\Delta t < 0$. Also existiert die Ableitung $\frac{\partial f}{\partial t}$. Durch leichte Umformung erhalten wir die *prospektive* Gleichung (8.9.7) (vgl. 8.7). Die *retrospektive* Gleichung kann man ebenfalls herleiten. Sie hat die Form

$$\frac{\partial f}{\partial \tau} = -m(\tau)\frac{\partial f}{\partial u} - \frac{1}{2}\sigma^2(\tau)\frac{\partial^2 f}{\partial u^2}. \tag{8.9.18}$$

Bachelier [1] fand die Gleichungen (8.9.7) und (8.9.18), gab aber keinen exakten Beweis an. Die Gleichungen stellen einen Spezialfall von Gleichungen dar, die von Kolmogoroff in [3] für rein stetige Prozesse hergeleitet wurden.

Um die Dichtefunktion $f(\tau, t, u)$ zu bestimmen, setzen wir

$$\begin{aligned} v &= u - \int_\tau^t m(\alpha)\,d\alpha, \\ t^* &= \int_\tau^t \sigma^2(\alpha)\,d\alpha \end{aligned} \tag{8.9.19}$$

und bezeichnen mit $g(\tau, t^*, v)$ die neue Dichtefunktion. Dann erhalten wir wegen $\frac{\partial u}{\partial v} = 1$ nach (2.4.10)

$$g(\tau, t^*, v) = f[\tau, t(t^*), u(v, t^*)],$$

wobei $t(t^*)$ und $u(v, t^*)$ die Umkehrung der Transformation (8.9.19) ist. Aus der letzten Gleichung und der Formel (8.9.7) ergibt sich

$$\frac{\partial g}{\partial t^*} = \frac{1}{2}\frac{\partial^2 g}{\partial v^2}. \tag{8.9.20}$$

Wenn wir ganz analog vorgehen, liefert (8.9.18)

$$\frac{\partial g}{\partial \tau^*} = -\frac{1}{2}\frac{\partial^2 g}{\partial v^2}. \tag{8.9.21}$$

Wir haben die Wärmeleitungsgleichungen erhalten. Die einzige Lösung dieser Gleichungen, die eine Dichtefunktion ist und (8.9.1) bis (8.9.6) erfüllt, ist die Funktion (vgl. COURANT und HILBERT [1])

$$g(\tau, t^*, v) = \frac{1}{\sqrt{2\pi t^*}} \exp\left(-\frac{v^2}{2t^*}\right). \tag{8.9.22}$$

Wenn wir

$$M(\tau, t) = \int_\tau^t m(\alpha)\, d\alpha, \qquad B^2(\tau, t) = t^*$$

setzen, erhalten wir

$$f(\tau, t, u) = \frac{1}{B(\tau, t)\sqrt{2\pi}} \exp\left\{-\frac{[u - M(\tau, t)]^2}{2B^2(\tau, t)}\right\}. \tag{8.9.23}$$

Der Zuwachs $X_t - X_\tau$ ist also normal verteilt und hat den Mittelwert $M(\tau, t)$ und die Standardabweichung $B(\tau, t)$. Da für beliebige

$$t_1, t_2, \ldots, t_n \qquad (0 \leqq t_1 < t_2 < \cdots < t_n < \infty)$$

die Gleichung

$$B^2(t_1, t_n) = \sum_{k=1}^{n-1} B^2(t_k, t_{k+1})$$

erfüllt ist, hat der betrachtete Prozeß unabhängige Zuwächse.

Wir wollen jetzt den folgenden Spezialfall eines solchen Prozesses betrachten: Die Zuwächse seien unabhängig und homogen, und der Mittelwert der Zuwächse sei identisch gleich Null. Es ist also

$$M(\tau, t) \equiv 0, \qquad B^2(\tau, t) \equiv (t - \tau)\,\sigma^2$$

Die Formel (8.9.23) hat hier die Gestalt

$$f(\tau, t, u) = \frac{1}{\sigma\sqrt{2\pi(t-\tau)}} \exp\left[-\frac{u^2}{2(t-\tau)\sigma^2}\right]. \tag{8.9.24}$$

Wir wählen insbesondere $\tau = 0$ und setzen voraus, daß $P(X_0 = 0) = 1$ ist. Dann ist X_t der Zuwachs in dem Intervall $[0, t)$. Die Dichtefunktion $f(t, u)$ der

Zufallsvariablen X_t hat die Gestalt

$$f(t, u) = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left(-\frac{u^2}{2t\sigma^2}\right). \tag{8.9.25}$$

Definition 8.9.1. Einen Prozeß $\{X_t, 0 < t < \infty\}$, der der Gleichung $P(X_0 = 0) = 1$ genügt, homogene und unabhängige Zuwächse hat und dessen Zufallsvariable X_t die Dichtefunktion (8.9.25) besitzen, nennt man *Brownschen Bewegungsprozeß* oder *Wienerschen Prozeß*.

Woher diese Namen kommen, wollen wir im weiteren Verlauf dieses Kapitels erklären.

Für den Wienerschen Prozeß ist also

$$\begin{aligned} &1. \quad P(X_0 = 0) = 1, \\ &2. \quad E(X_t - X_\tau) = 0, \\ &3. \quad D^2(X_t - X_\tau) = (t - \tau)\sigma^2. \end{aligned} \tag{8.9.26}$$

wobei τ und t $(0 \leqq \tau < t)$ beliebig sind. Weiter ist für beliebige $n = 2, 3, 4, \ldots$ und beliebige Parameterwerte t_i $(i = 1, 2, \ldots, n)$, wobei $t_1 < \cdots < t_n$ ist,

$$X_{t_n} = X_{t_1} + (X_{t_2} - X_{t_1}) + \cdots + (X_{t_n} - X_{t_{n-1}}).$$

Die Zufallsvariablen $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ sind unabhängig und normal verteilt. Die Zufallsvariablen $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ haben als lineare Funktionen normal verteilter unabhängiger Zufallsvariabler nach Satz 5.11.1 eine gemeinsame Normalverteilung. Dabei ist

$$E(X_{t_i}) = 0, \quad D^2(X_{t_i}) = t_i\sigma^2 \qquad (i = 1, \ldots, n),$$

$$E(X_{t_i}X_{t_j}) = t_i\sigma^2 \qquad (i, j = 1, 2, \ldots, n; \quad i < j).$$

Die Matrix M der Momente zweiter Ordnung der Zufallsvariablen

$$(X_{t_1}, \ldots, X_{t_n})$$

hat die Form

$$\frac{M}{\sigma^2} = \begin{pmatrix} t_1 & t_1 & \ldots & t_1 \\ t_1 & t_2 & \ldots & t_2 \\ \cdot & \cdot & \cdots & \cdot \\ t_1 & t_2 & \ldots & t_n \end{pmatrix}.$$

Beispiel 8.9.1 (Brownsche Bewegung). In einer Flüssigkeit führt ein mikroskopisches Teilchen unter dem Einfluß der Stöße der Flüssigkeitsmoleküle Bewegungen aus. Diese Erscheinung beobachtete Brown im Jahre 1826. Wir betrachten die Bewegung eines solchen

Teilchens und bezeichnen mit X_t seine Abszisse im Augenblick t. Wie EINSTEIN [1] und SMOLUCHOWSKI [1] zeigten, wird eine gute Annäherung erzielt, wenn man X_t als Brownschen Bewegungsprozeß auffaßt. Das erklärt sich im wesentlichen daraus, daß Kontakte zwischen dem untersuchten Teilchen und den Flüssigkeitsmolekülen nur im Augenblick des Zusammenstoßes auftreten. Die Stöße erfolgen unregelmäßig und häufig. Ist die Differenz $t - \tau$ im Verhältnis zur Zeit, die zwischen zwei aufeinanderfolgenden Stößen vergeht, groß, dann kann der Zuwachs $X_t - X_\tau$ als Summe vieler kleiner Zuwächse betrachtet werden. Befindet sich die Flüssigkeit im makroskopischen Gleichgewicht, dann kann man annehmen, daß erstens die Zuwächse nur von der Länge des Zeitabschnittes abhängen, also homogen sind, und zweitens Zuwächse in disjunkten Intervallen unabhängig sind. Weiter nehmen wir an, daß die Bewegung symmetrisch ist, also $E(X_t - X_\tau) \equiv 0$. EINSTEIN zeigte, daß

$$E[(X_t - X_\tau)^2] = 2D^2 t$$

ist, wobei D die Diffusionskonstante der Flüssigkeit ist. Also ist $\sigma^2 = 2D^2$ eine für die Flüssigkeit charakteristische Größe. Wir können alle Voraussetzungen, die zum Brownschen Bewegungsprozeß führten, als erfüllt ansehen und annehmen, daß ein derartiger Prozeß vorliegt, was auch durch Beobachtungen bestätigt wird.

Wie wir schon oben ausführten, können wir $X_t - X_\tau$ als Summe von vielen unabhängigen Zufallsvariablen betrachten. Auf Grund des zentralen Grenzwertsatzes liegt es also auch nahe zu vermuten, daß $X_t - X_\tau$ eine Normalverteilung hat.

Die Bewegung von Gasmolekülen bei niedrigem Druck kann ebenfalls als Brownsche Bewegung angesehen werden. Allgemein kann als Brownsche Bewegung die Bewegung eines Körpers in einem Medium betrachtet werden, wenn die Ausdehnungen dieses Körpers mikroskopisch klein im Verhältnis zu den Ausdehnungen des Mediums und zu den anderen Elementen sind, die sich darin befinden, und wenn Berührungen dieses Körpers mit den Elementen des Mediums plötzlich erfolgen und vom Zufall gesteuert sind. So können wir zum Beispiel als Körper einen „kleinen" Stern betrachten, d. h. einen Stern, dessen Ausdehnungen klein sind im Verhältnis zu den anderen Sternen, die sich im kosmischen Raum befinden. Hierbei spielt der kosmische Raum die Rolle der Flüssigkeit, in der die Brownsche Bewegung stattfindet. Unter einem „Stoß" verstehen wir, daß sich der Stern im Kräftefeld anderer Sterne befindet. Die Bahn eines solchen „kleinen" Sterns ist vom Zufall abhängig, und zu ihrer Beschreibung kann man das Modell eines Brownschen Bewegungsprozesses heranziehen.

Beispiel 8.9.2. In Beispiel 5.11.1 betrachteten wir die Geschwindigkeit eines Moleküls in einem idealen Gas und stellten fest, daß jede Komponente des Geschwindigkeitsvektors normal verteilt ist. Zu demselben Ergebnis kommt man, wenn man das Modell des Brownschen Bewegungsprozesses anwendet. Die Durchführung der entsprechenden Überlegung überlassen wir dem Leser.

Zum Abschluß des Paragraphen bemerken wir noch, daß die von uns betrachteten Prozesse mit unabhängigen und homogenen Zuwächsen, nämlich der Poissonsche Prozeß, der Geburtsprozeß und der Wienersche Prozeß, Spezialfälle der umfangreichen Klasse der Prozesse mit unabhängigen und homogenen Zuwächsen sind. Hier begnügen wir uns damit, den Leser auf die Arbeiten von CRAMÉR [1], DE FINETTI [1], LÉVY [2] und KOLMOGOROFF [5] zu verweisen.

Es sei erwähnt, daß viele Verallgemeinerungen des Prozesses der Brownschen

Bewegung (auf den mehrdimensionalen Fall und auf andere Fälle) untersucht worden sind. Hier seien nur zwei Arbeiten genannt, und zwar die von LÉVY [11] und die von ITÔ [1].

8.10. Stationäre Prozesse

A. Jetzt wollen wir uns kurz mit einer Klasse von Prozessen beschäftigen, die man als *stationäre Prozesse* bezeichnet. Die Theorie dieser Prozesse hat sich in der jüngsten Zeit entwickelt und fand immer zahlreichere Anwendungen. Zunächst wollen wir einige allgemeine Bemerkungen machen. Stationäre Prozesse haben die Eigenschaft, daß die Wahrscheinlichkeitsverteilung oder zumindest manche Momente der Zufallsvariablen des Prozesses von einer Verschiebung des Anfangspunktes auf der Zeitachse t unabhängig sind. Die Klassen der Markoffschen Prozesse und der stationären Prozesse schließen sich nicht gegenseitig aus. Manche Markoffschen Prozesse sind stationäre Prozesse, obwohl andererseits viele Zufallserscheinungen in der Industrie, in der Meteorologie, in der Ökonomie und in anderen Gebieten dem Modell des Markoffschen Prozesses nicht genügen, sich aber dafür mit Hilfe der Theorie der stationären Prozesse erfassen lassen.

B. Im folgenden werden wir komplexe Zufallsvariable betrachten müssen, die wir nunmehr definieren wollen.

Definition 8.10.1. Es seien X und Y reelle Zufallsvariable im Sinne der Definition 2.1.2. Dann bezeichnet man mit $Z = X + iY$ (i ist die imaginäre Einheit) eine *komplexe Zufallsvariable*.

Die Verteilung der Zufallsvariablen Z wird durch die zweidimensionale Verteilungsfunktion $F(x, y) = P(X < x, Y < y)$ bestimmt.

Definition 8.10.1′. Es seien $\{X_t, t \in I\}$ und $\{Y_t, t \in I\}$ reelle stochastische Prozesse im Sinne der Definition 8.1.1, und es sei $Z_t \equiv X_t + iY_t$ $(t \in I)$. Dann nennt man $\{Z_t, t \in I\}$ einen *komplexen stochastischen Prozeß*.

Definition 8.10.2. Den Ausdruck

$$D^2(Z) = E(|Z - E(Z)|^2)$$

nennt man die *Varianz der komplexen Zufallsvariablen Z*.

Wie der Leser selbst verifizieren kann, gilt

$$D^2(Z) = E(|Z|^2) - |E(Z)|^2 = D^2(X) + D^2(Y).$$

Definition 8.10.3. Den Ausdruck

$$E\left\{[Z_1 - E(Z_1)]\left[\overline{Z_2 - E(Z_2)}\right]\right\}$$

nennt man die *Kovarianz der komplexen Zufallsvariablen* Z_1 *und* Z_2, der Komponenten des Vektors (Z_1, Z_2).

Es sei hier darauf hingewiesen, daß im Gegensatz zum reellen Fall die Kovarianz der Komponenten des Vektors (Z_1, Z_2) nicht gleich der Kovarianz des Vektors (Z_2, Z_1) ist.

C. Wir wenden uns nun der Definition eines stationären Prozesses zu.

Definition 8.10.4. Ein stochastischer komplexer Prozeß $\{Z_t, -\infty < t < \infty\}$ heißt *stationär im engeren Sinne*, wenn für $n = 1, 2, \ldots$ und für beliebige reelle $t_1, \ldots, t_n$ und τ die Gleichung

$$\begin{aligned} &F_{t_1, \ldots, t_n}(x_1, \ldots, x_n, y_1, \ldots, y_n) \\ &= P(X_{t_1} < x_1, \ldots, X_{t_n} < x_n, Y_{t_1} < y_1, \ldots, Y_{t_n} < y_n) \\ &= P(X_{t_1+\tau} < x_1, \ldots, X_{t_n+\tau} < x_n, Y_{t_1+\tau} < y_1, \ldots, Y_{t_n+\tau} < y_n) \\ &= F_{t_1+\tau, \ldots, t_n+\tau}(x_1, \ldots, x_n, y_1, \ldots, y_n) \end{aligned} \tag{8.10.1}$$

erfüllt ist.

Ein reeller Prozeß $\{X_t, -\infty < t < \infty\}$ heißt *stationär im engeren Sinne*, wenn für die betrachteten $n, t_1, \ldots, t_n$ und τ die Gleichung

$$F_{t_1, \ldots, t_n}(x_1, \ldots, x_n) = F_{t_1+\tau, \ldots, t_n+\tau}(x_1, \ldots, x_n) \tag{8.10.1'}$$

gilt.

Aus dieser Definition folgt insbesondere, daß $F_t(x, y) = F_{t-t}(x, y) = F_0(x,y)$ für jedes reelle t und daß $F_t(x) = F_0(x)$ für einen reellen Prozeß gilt. Somit haben alle Z_t (im reellen Fall alle X_t) die gleiche Verteilung.

Alle zweidimensionalen Verteilungsfunktionen sind nur von der Differenz der Werte des Parameters t abhängig, und zwar gilt

$$F_{t_1,t_2}(x_1, x_2, y_1, y_2) = F_{0,\, t_2-t_1}(x_1, x_2, y_1, y_2)$$

bzw. für reelle stationäre Prozesse

$$F_{t_1,t_2}(x_1, x_2) = F_{0,\, t_2-t_1}(x_1, x_2).$$

Wenn die Momente zweiter Ordnung der Zufallsvariablen X_t und Y_t existieren und $Z_t = X_t + i\,Y_t$ ein stationärer Prozeß im engeren Sinne ist, dann gelten wegen (8.10.1) für alle reellen t und τ die Gleichungen

$$\begin{aligned} &E(Z_t) = E(X_t + i\,Y_t) = m_1 + i m_2 = m = \text{const}, \\ &D^2(Z_t) = D^2(X_t) + D^2(Y_t) = \sigma^2 = \text{const}, \\ &E\left\{[Z_{t+\tau} - E(Z_{t+\tau})]\left[\overline{Z_t - E(Z_t)}\right]\right\} = R(\tau). \end{aligned} \tag{8.10.2}$$

Die Funktion $R(\tau)$ heißt *Korrelationsfunktion*. Wie man leicht sieht, gilt $R(0) = D^2(Z_t)$ und $R(-\tau) = \overline{R(\tau)}$. Für reelle Prozesse gehen die Formeln (8.10.2) über in

$$\begin{aligned} &E(X_t) = m = \text{const}, \\ &D^2(X_t) = \sigma^2 = \text{const}, \qquad (8.10.2') \\ &E\{[X_{t+\tau} - E(X_{t+\tau})][X_t - E(X_t)]\} = R(\tau). \end{aligned}$$

Liegt ein reeller Prozeß vor und ist hierbei $m = 0$ und $\sigma^2 = 1$, so ist $R(\tau)$ einfach der *Korrelationskoeffizient* der Zufallsvariablen X_t und $X_{t+\tau}$.

Beispiel 8.10.1. Es bezeichne X_t die Temperatur eines festgewählten Punktes in der Atmosphäre zur Zeit t. Wir betrachten die Temperaturschwankungen in kleinen Zeitabschnitten, so daß es sich um sehr kleine Temperaturschwankungen turbulenten Charakters handelt. Diese Schwankungen verursachen senkrechte Wärmebewegungen und haben bedeutenden Einfluß auf den Verlauf vieler meteorologischer Erscheinungen. Man kann nicht annehmen, daß X_t ein Markoffscher Prozeß ist, da die Temperatur des Punktes zur Zeit t_0 nicht die Wahrscheinlichkeitsverteilung der Zufallsvariablen X_t für $t > t_0$ bestimmt. Diese Verteilung hängt auch von der Temperatur in früheren Augenblicken ab; denn die Intensität der lokalen Temperaturschwankung ist z. B. in den einzelnen Jahreszeiten verschieden. Aber man kann bei beständigem Wetter annehmen, daß X_t ein stationärer Prozeß ist. Die Beobachtungen der lokalen atmosphärischen Temperaturschwankungen wurden mit einem Mikrothermometer durchgeführt, dessen Konstruktion KRETSCHMER [1] angegeben hat. Wir haben die beobachtete Realisierung des Prozesses X_t in Abb. 8.10.1 dargestellt.

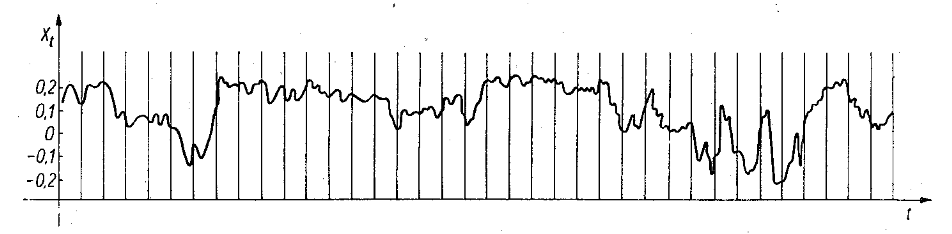


Abb. 8.10.1

Besonders wichtig ist der *normale stationäre Prozeß*, d. h. ein stationärer Prozeß, der den Beziehungen (8.10.2') mit $m = 0$ und $\sigma = 1$ genügt und der die Eigenschaft hat, daß für $n = 1, 2, 3, \ldots$ und beliebige $t_m (m = 1, 2, \ldots, n)$ die Zufallsvariable $(X_{t_1}, \ldots, X_{t_n})$ normal verteilt ist. Der Formel (5.11.9) entnimmt man, daß $(X_{t_1}, \ldots, X_{t_n})$ dann die charakteristische Funktion

$$\varphi_{t_1, \ldots, t_n}(v_1, \ldots, v_n) = \exp\left[-\frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} R(t_k - t_j) v_j v_k\right]$$

hat.

In den Anwendungen ist es gewöhnlich schwierig, mehrdimensionale Verteilungen der Zufallsvariablen X_t zu untersuchen. Vielfach kann man aus dem Beobachtungsmaterial nur die Momente der Zufallsvariablen berechnen. Diesem Tatbestand wollen wir die Begriffsbildung eines stationären Prozesses im weiteren Sinne anpassen.

Definition 8.10.5. Ein komplexer Prozeß $\{Z_t, -\infty < t < \infty\}$ heißt ein *stationärer Prozeß im weiteren Sinne* (oder *im Chintschinschen Sinne*), wenn die Erwartungswerte, Varianzen und Kovarianzen der Zufallsvariablen dieses Prozesses existieren und den Gleichungen (8.10.2) genügen. Ein reeller Prozeß $\{X_t, -\infty < t < \infty\}$ heißt *stationär im weiteren Sinne*, wenn seine Zufallsvariablen endliche Momente erster und zweiter Ordnung besitzen und den Beziehungen (8.10.2′) genügen.

Aus dieser Definition geht hervor, daß ein stationärer Prozeß im engeren Sinne, der Momente zweiter Ordnung besitzt, ein stationärer Prozeß im weiteren Sinne ist. Offenbar ist der in diesem Abschnitt betrachtete reelle normale Prozeß stationär sowohl im engeren als auch im weiteren Sinne.

D. Wir wollen nun den Begriff der Stetigkeit eines stationären Prozesses einführen.

Definition 8.10.6. Ein Prozeß $\{Z_t, -\infty < t < \infty\}$, der stationär im weiteren Sinne ist, heißt *stetig*, wenn für jedes reelle t

$$\lim_{\tau\to 0} E(|Z_{t+\tau} - Z_t|^2) = 0 \tag{8.10.3}$$

gilt.

Satz 8.10.1. *Eine notwendige und hinreichende Bedingung dafür, daß ein im weiteren Sinne stationärer Prozeß stetig ist, ist die Stetigkeit seiner Korrelationsfunktion $R(\tau)$ an der Stelle $\tau = 0$.*

Beweis. Der im weiteren Sinne stationäre Prozeß $\{Z_t, -\infty < t < \infty\}$ sei stetig. Unter Berücksichtigung der Gleichungen (8.10.2) erhalten wir nach einfachen Umformungen

$$\begin{aligned} |R(\tau + \Delta\tau) - R(\tau)| &= \left|E\left\{(Z_{\tau+\Delta\tau} - Z_\tau)\left[\overline{Z_0 - E(Z_0)}\right]\right\}\right| \\ &\leqq E\left\{\left|(Z_{\tau+\Delta\tau} - Z_\tau)\left[\overline{Z_0 - E(Z_0)}\right]\right|\right\}. \end{aligned} \tag{8.10.4}$$

Auf Grund der Schwarzschen Ungleichung (vgl. etwa Fichtenholz [1]) erhalten wir

$$\begin{aligned} &E\left\{\left|(Z_{\tau+\Delta\tau} - Z_\tau)\left[\overline{Z_0 - E(Z_0)}\right]\right|\right\} \\ &\leqq \sqrt{E(|Z_{\tau+\tau\Delta} - Z_\tau|^2)E[|Z_0 - E(Z_0)|^2]}. \end{aligned} \tag{8.10.5}$$

Wegen $D^2(Z_0) < \infty$ liefern die Beziehungen (8.10.3), (8.10.4) und (8.10.5)

$$\lim_{\Delta\tau\to 0} |R(\tau + \Delta\tau) - R(\tau)| = 0, \tag{8.10.6}$$

und die Funktion $R(\tau)$ ist somit stetig.

Nun sei $R(\tau)$ stetig an der Stelle $\tau = 0$. Da es sich um einen stationären Prozeß handelt, haben wir

$$
\begin{aligned}
E(|Z_{\tau+\tau} - Z_t|^2) &= E\{|Z_\tau - E(Z_\tau) - [Z_0 - E(Z_0)]|^2\} \\
&= E[|Z_\tau - E(Z_\tau)|^2] + E[|Z_0 - E(Z_0)|^2] \\
&\quad - E\left\{[Z_0 - E(Z_0)]\left[\overline{Z_\tau - E(Z_\tau)}\right]\right\} \\
&\quad - E\left\{[Z_\tau - E(Z_\tau)]\left[\overline{Z_0 - E(Z_0)}\right]\right\} \\
&= 2D^2(Z_0) - R(\tau) - R(-\tau).
\end{aligned}
$$

Wegen $R(0) = D^2(Z_0)$ und der vorausgesetzten Stetigkeit der Funktion $R(\tau)$ an der Stelle $\tau = 0$ erhalten wir aus den letzten Gleichungen

$$\lim_{\tau \to 0} E(|Z_{t+\tau} - Z_t|^2) = 0,$$

und der betrachtete Prozeß ist somit stetig.

Für das weitere wollen wir annehmen, daß in den Gleichungen (8.10.2) und (8.10.2′) stets

$$m = 0, \qquad \sigma = 1 \tag{8.10.7}$$

gilt.

Nun wollen wir einen von CHINTSCHIN [3] bewiesenen Satz kennenlernen, der für die Theorie der stationären Prozesse von fundamentaler Bedeutung ist.

Satz 8.10.2. *Eine notwendige und hinreichende Bedingung dafür, daß* $R(\tau)$ *Korrelationsfunktion eines im weiteren Sinne stationären stetigen Prozesses* $\{Z_t, -\infty < t < \infty\}$ *ist, der den Beziehungen* (8.10.7) *genügt, ist die Existenz einer Verteilungsfunktion F mit der Eigenschaft, daß*

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\lambda\tau}\, dF(\lambda) \tag{8.10.8}$$

gilt.

Beweis. Es sei $R(\tau)$ Korrelationsfunktion des betrachteten stationären Prozesses. Aus (8.10.2) und (8.10.7) folgt dann

$$R(0) = 1. \tag{8.10.9}$$

Ferner haben wir für $n = 1, 2, \ldots$ und beliebige reelle $t_1, \ldots, t_n$ sowie komplexe $a_1, \ldots, a_n$ die Beziehung

$$
\begin{aligned}
\sum_{j,k=1}^{n} R(t_j - t_k) a_j \overline{a_k} &= \sum_{j,k=1}^{n} a_j \overline{a_k} E\left(Z_{t_j - t_k} \overline{Z_0}\right) \\
&= \sum_{j,k=1}^{n} a_j \overline{a_k} E\left(Z_{t_j} \overline{Z_{t_k}}\right) = E\left(\left|\sum_{j=1}^{n} a_j Z_{t_j}\right|^2\right) \geqq 0,
\end{aligned}
$$

und somit ist $R(\tau)$ eine positiv definite Funktion. Beachten wir schließlich (8.10.9) sowie die Tatsache, daß $R(\tau)$ auf Grund des Satzes 8.10.1 stetig ist, so folgt nach Satz 4.1.1 die Existenz einer Verteilungsfunktion $F(\lambda)$, die der Beziehung (8.10.8) genügt.

Nun möge $R(\tau)$ der Beziehung (8.10.8) genügen, wobei $F(\lambda)$ eine Verteilungsfunktion bedeutet. Wir müssen die Existenz eines im weiteren Sinne stationären stetigen Prozesses $\{Z_t, -\infty < t < \infty\}$ nachweisen, der den Gleichungen (8.10.7) genügt und dessen Korrelationsfunktion die gegebene Funktion $R(\tau)$ ist. Zum Beweis genügt es offenbar, ein Beispiel eines solchen Prozesses anzugeben. Der Einfachheit halber wollen wir uns mit dem Fall begnügen, daß $R(\tau)$ eine reelle Funktion ist. Ein Beispiel für einen solchen Prozeß ist der reelle stationäre normale Prozeß, den wir in 8.10.C betrachtet hatten. Sämtliche endlichdimensionalen Verteilungen dieses Prozesses, d. h. Verteilungen der Vektoren $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ für beliebige n und $t_1, t_2, \ldots, t_n$ sind durch die Matrix der Momente zweiter Ordnung bestimmt, die ihrerseits durch die Korrelationsfunktion $R(\tau)$ eindeutig bestimmt sind.

Damit ist der Satz bewiesen.

Der Leser wird bereits festgestellt haben, daß bei einer Normierung gemäß (8.10.7) die Korrelationsfunktion $R(\tau)$ auf Grund des Satzes 8.10.2 sämtliche Eigenschaften einer charakteristischen Funktion besitzt.

Die in (8.10.8) auftretende Funktion $F(\lambda)$ führt die Bezeichnung *spektrale Verteilungsfunktion*. Aus dem Satz 4.5.1 folgt, daß die Korrelationsfunktion $R(\tau)$ die spektrale Verteilungsfunktion $F(\lambda)$ eindeutig bestimmt.

Existiert eine Dichtefunktion $f(\lambda)$, so heißt der zugehörige stationäre Prozeß ein *Prozeß mit stetigem Spektrum*. Falls hierbei $R(\tau)$ eine absolut differenzierbare Funktion auf der gesamten τ-Achse ist $(-\infty < \tau < \infty)$, läßt sich die Spektraldichte $f(\lambda)$ aus der Beziehung (4.5.6) ermitteln, die nunmehr die Gestalt

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda\tau} R(\tau)\, d\tau$$

annimmt.

Weist die spektrale Verteilungsfunktion $F(\lambda)$ Zunahmen nur an den Sprungstellen λ_k auf, so bezeichnet man den zugehörigen Prozeß als einen *Prozeß mit diskretem Spektrum*. Wir sagen, daß (8.10.8) eine *Spektraldarstellung* der Korrelationsfunktion liefert.

E. Wir betrachten den durch

$$Z_t = \sum_{k=1}^{n} W_k\, e^{i\lambda_k t} \tag{8.10.10}$$

definierten Prozeß $\{Z_t, -\infty < t < \infty\}$, wobei die λ_k reelle Konstanten und die $W_k (k = 1, 2, \ldots, n)$ komplexe Zufallsvariable sind, die den Gleichungen

$$\begin{aligned} &E(W_k) = 0, \quad E(|W_k|^2) = \sigma_k^2, \\ &E(W_k \overline{W_j}) = 0 \qquad (k \neq j) \end{aligned} \tag{8.10.11}$$

genügen; hierbei gilt $\sigma_k^2 > 0$ $(k = 1, \ldots, n)$ und $\sum_{k=1}^{n} \sigma_k^2 = 1$. Die letzte der Gleichungen (8.10.11) besagt, daß die Variablen W_k nicht korreliert sind.

Der Prozeß $\{Z_t, -\infty < t < \infty\}$, der durch die Gleichung (8.10.10) definiert ist, stellt eine Überlagerung von n periodischen Schwingungen mit zufälliger Amplitude und zufälliger Phase dar. Wir setzen $W_k = U_k + i V_k$, wobei U_k und V_k reell sind. Der Realteil des Ausdrucks $W_k e^{i\lambda_k t}$ ergibt sich dann zu

$$U_k \cos \lambda_k t - V_k \sin \lambda_k t = R_k \sin (\lambda_k t + \alpha_k)$$

mit

$$R_k = \sqrt{U_k^2 + V_k^2}, \quad \sin \alpha_k = \frac{U_k}{R_k}, \quad \cos \alpha_k = -\frac{V_k}{R_k}.$$

Das ist eine Sinuskurve mit der konstanten Frequenz λ_k, der zufälligen Amplitude R_k und der zufälligen Phase α_k. Somit ist der Realteil des Prozesses Z_t eine Überlagerung solcher Sinuskurven.

Der durch (8.10.10) definierte Prozeß Z_t ist ein stationärer stetiger Prozeß im weiteren Sinne und genügt den Gleichungen (8.10.7). In der Tat, wegen (8.10.10) und (8.10.11) erhalten wir

$$\begin{aligned} E(Z_t) &= 0, \\ D^2(Z_t) &= E(|Z_t|^2) = \sum_{k=1}^{n} \sigma_k^2 = 1, \\ R(\tau) &= E\left(Z_{t+\tau} \overline{Z_t}\right) \\ &= E\left\{\left[\sum_{k=1}^{n} W_k e^{i\lambda_k (t+\tau)}\right]\left[\sum_{k=1}^{n} \overline{W_k} e^{-i\lambda_k t}\right]\right\} \\ &= E\left(\sum_{k=1}^{n} W_k \overline{W_k} e^{i\lambda_k \tau}\right) = \sum_{k=1}^{n} \sigma_k^2 e^{i\lambda_k \tau}. \end{aligned} \tag{8.10.12}$$

Damit ist gezeigt, daß der Prozeß ein stationärer Prozeß im weiteren Sinne ist und den Gleichungen (8.10.7) genügt. Seine Stetigkeit folgt aus der Stetigkeit der Funktion $R(\tau)$.

Die Verteilungsfunktion $F(\lambda)$, die gemäß (8.10.8) der durch (8.10.12) gegebenen Korrelationsfunktion $R(\tau)$ entspricht, ist eine stückweise konstante Funktion, die nur in den Punkten λ_k mit den jeweiligen Sprunghöhen σ_k^2 anwächst. Der betrachtete Prozeß hat somit ein diskretes Spektrum.

Es läßt sich leicht zeigen, daß ein reeller Prozeß mit einer Korrelationsfunktion der Form

$$R(\tau) = 2\sum_{k=1}^{m}\sigma_k^2 \cos \lambda_k \tau$$

vorliegt, wenn $n = 2m$ ist und hierbei $2m$ Summanden der Summe (8.10.10) m konjugiert komplexe Paare bilden.

F. Der durch (8.10.10) definierte Prozeß bildet den Ausgangspunkt für die harmonische Analyse der stationären stochastischen Prozesse. Um wenigstens eine grobe Information über das Wesen dieser Problemstellung zu vermitteln, müssen wir eine neue Art der Konvergenz einführen, die mit dem in der Definition 8.10.6 gegebenen Stetigkeitsbegriff in Zusammenhang steht.

Definition 8.10.7. Eine Folge (komplexer oder reeller) Zufallsvariabler $\{U_n\}$, die Momente zweiter Ordnung besitzen, nennt man *im quadratischen Mittel konvergent gegen die Zufallsvariable* U, wenn die Bedingung

$$\lim_{n\to\infty} E(|U_n - U|^2) = 0 \tag{8.10.13}$$

erfüllt ist.

Wir betrachten nun den durch die Formel

$$Z_t = \lim_{n\to\infty} \sum_{k=1}^{n} W_k e^{i\lambda_k t} = \sum_{k=1}^{\infty} W_k e^{i\lambda_k t} \tag{8.10.14}$$

definierten Prozeß $\{Z_t, -\infty < t < \infty\}$, wobei die W_k $(k = 1, 2, \ldots)$ komplexe Zufallsvariable sind, die den Gleichungen (8.10.11) und darüber hinaus der Gleichung

$$\sum_{k=1}^{\infty}\sigma_k^2 = 1 \tag{8.10.15}$$

genügen. Der Grenzübergang in (8.10.14) ist hierbei im Sinne der Definition 8.10.7 zu verstehen. Es läßt sich zeigen, daß dieser Grenzwert existiert (siehe etwa Doob [5], Kap. IV, § 4).

Der Prozeß Z_t ist stationär im weiteren Sinne, und seine Korrelationsfunktion $R(\tau)$ hat die Form

$$R(\tau) = \sum_{k=1}^{\infty}\sigma_k^2 e^{i\lambda_k \tau}. \tag{8.10.16}$$

Slutski [3] hat gezeigt, daß jeder stetige, im weiteren Sinne stationäre Prozeß mit diskretem Spektrum sich in Form der auf der rechten Seite von (8.10.14) auftretenden Reihe darstellen läßt. Das ist ein Spezialfall des *Satzes über die*

Spektraldarstellung stationärer Prozesse. Einen allgemeineren Satz über die Spektraldarstellung stationärer Prozesse, der Prozesse mit kontinuierlichem Spektrum umfaßt und in dem die auf der rechten Seite von (8.10.14) stehende Reihe durch eigens für diesen Zweck definierte Integrale zu ersetzen sind, hat KOLMOGOROFF bewiesen [10], [11]. Den Satz für den in 8.10.C betrachteten normalen stationären Prozeß haben KAC und SIEGERT [1] formuliert; sie haben eine effektive Formel gefunden, die X_t in Form einer Reihe darstellt. Sätze über die Spektraldarstellung von Prozessen (ohne die Voraussetzung, daß diese stationär sind), deren Zufallsvariable endliche Momente zweiter Ordnung besitzen, haben LOÈVE [1], [2], KARHUNEN [1], CRAMÉR [6] und GETOOR [1] angegeben.

G. Wir wollen nun einige Betrachtungen der Interpretation einer spektralen Verteilungsfunktion in physikalischen Problemen widmen. Es sei Z_t der reelle stationäre Prozeß, den wir am Schluß von 8.10.E betrachtet hatten. Es sei etwa Z_t die Spannungsschwankung eines elektrischen Stromes, die sich als Summe von Sinuskurven mit konstanten Frequenzen, jedoch mit zufälligen Phasen und Amplituden darstellen läßt. Dann stellt die spektrale Verteilungsfunktion dieses Prozesses die Verteilung der elektrischen Energie dar, die von den Komponenten mit den vorgegebenen Frequenzen mitgeführt wird, vorausgesetzt, daß die Gesamtenergie gleich 1 gesetzt wird. Nun seien λ' und λ'' $(\lambda' < \lambda'')$ beliebige Punkte auf der λ-Achse. Dann gilt

$$F(\lambda'') - F(\lambda') = \sum \frac{1}{2}\, \sigma_k^2,$$

wobei die Summation über alle λ_k des Intervalls $[\lambda', \lambda'')$ erstreckt wird. Der Zuwachs der spektralen Verteilungsfunktion im Intervall $[\lambda', \lambda'')$ ist gleich der mittleren Energie der Komponenten des Prozesses, deren Schwingungsfrequenzen zwischen λ' und λ'' liegen. Für stationäre Prozesse mit kontinuierlichem Spektrum ist die Interpretation analog, doch sind wir hier nicht in der Lage, diese zu formulieren. Wir verweisen den Leser auf die Arbeit von JAGLOM [1] bzw. auf das Buch von GRENANDER und ROSENBLATT [1].

H. Neben den stationären Prozessen mit stetiger Zeit betrachtet man auch stationäre Prozesse mit diskreter Zeit, d. h. stationäre Folgen von Zufallsvariablen. Die Definitionen 8.10.4 und 8.10.5 sind in diesem Fall die gleichen, mit dem einen Unterschied, daß t und τ nun die ganzen Zahlen im Intervall $(-\infty, \infty)$ durchlaufen. Der Begriff der Stetigkeit eines stationären Prozesses ist jetzt überflüssig, denn τ kann nur ganzzahlige Werte annehmen.

Wir bemerken, daß bei nur ganzzahlige Werte durchlaufendem t für ganzzahliges j die Beziehung $\exp[i(\lambda + 2j\pi)t] = \exp(i\lambda t)$ gilt. Demnach können wir uns bei stationären Folgen darauf beschränken, nur die λ-Werte im Intervall $[-\pi, \pi]$ zu betrachten. Auch der Satz 8.10.2 gilt für stationäre Folgen, allerdings mit der Modifikation, daß die Integration in der Formel (8.10.2) über das Intervall $[-\pi, \pi]$ erstreckt wird. Der Beweis dafür stammt von HERGLOTZ [1].

Man findet ihn auch bei WOLD [1]. Desgleichen läßt sich auch der Satz von der Spektraldarstellung der Prozesse auf stationäre Folgen übertragen.

Beispiel 8.10.2. Es sei $\{X_t\}$ $(t = 0, \pm 1, \pm 2, \ldots)$ eine stationäre Folge (siehe Definition 7.5.3) von Zufallsvariablen mit den möglichen Zuständen x_i $(i = 1, 2, \ldots)$, die eine Markoffsche Kette bilden. Die Folge $\{X_t\}$ ist ein stationärer Prozeß im engeren Sinne mit diskreter Zeit. In der Tat, sind $t_1, t_2, \ldots, t_{n-1}, t_n$ und τ ganze Zahlen, so gilt wegen (7.2.3)

$$\begin{aligned}
&P(X_{t_1} = x_{i_1},\ X_{t_2} = x_{i_2}, \ldots,\ X_{t_{n-1}} = x_{i_{n-1}}, X_{t_n} = x_{t_n}) \\
&= P(X_{t_1} = x_{i_1})\, p_{i_1 i_2}(t_2 - t_1) \cdots p_{i_{n-1} i_n}(t_n - t_{n-1}) \\
&= P(X_{t_1+\tau} = x_{i_1}, X_{t_2+\tau} = x_{i_2}, \ldots, X_{t_{n-1}+\tau} = x_{i_{n-1}}, X_{t_n+\tau} = x_{i_n}).
\end{aligned}$$

Der Prozeß kann jedoch nicht stationär im weiteren Sinne sein, da wir keinerlei Voraussetzungen über die Momente der Zufallsvariablen X_t getroffen haben.

Beispiel 8.10.3. Es sei $\{X_t\}$ $(t = 0, \pm 1, \pm 2, \ldots)$ eine Folge reeller Zufallsvariabler mit den Erwartungswerten 0. Die Zufallsvariablen mögen paarweise nicht korreliert sein und die Standardabweichung 1 besitzen. Die Korrelationsfunktion $R(\tau)$ hat dann die Form

$$R(\tau) = E(X_{t+\tau} X_t) = \begin{cases} 1 & \text{für} \quad \tau = 0, \\ 0 & \text{für} \quad \tau \neq 0. \end{cases} \tag{8.10.17}$$

Wir bemerken, daß hier

$$R(\tau) = \int_{-\pi}^{\pi} \frac{\cos \lambda \tau}{2\pi}\, d\lambda$$

gilt und der betrachtete Prozeß somit die Spektraldichte

$$f(\lambda) = \frac{1}{2\pi} \tag{8.10.18}$$

hat.

Der in diesem Beispiel betrachtete stationäre Prozeß hat eine konstante Spektraldichte. Ein solcher Prozeß heißt *weißes Rauschen.* Diese Bezeichnung hängt mit der Interpretation der Spektralverteilung in physikalischen Problemen zusammen, von der in 8.10.G die Rede war. Sie beruht auf der Tatsache, daß in der Spektralverteilung des weißen Lichtes die Intensität sämtlicher Frequenzen die gleiche ist.

Die Frage der Gültigkeit des Gesetzes der großen Zahlen für eine im engeren Sinne stationäre Folge von Zufallsvariablen wurde in den Arbeiten von BIRKHOFF [1] und CHINTSCHIN [3] gelöst. IBRAGIMOW [1] hat ein weitreichendes Ergebnis bezüglich des zentralen Grenzwertsatzes für eine solche Folge gewonnen.

Für im weiteren Sinne stationäre Folgen von Zufallsvariablen läßt sich eine neue Art des Gesetzes der großen Zahlen betrachten, in dem die Konvergenz im Sinne der Definition 8.10.7 zu verstehen ist. Ein solches Gesetz der großen Zahlen findet der Leser z. B. in dem Buch von DOOB [5].

I. Ein Grundproblem in der Theorie der stationären Prozesse und Folgen im weiteren Sinne ist das *Extrapolationsproblem.* Es sei $\{X_t\}$ $(t = 0, \pm 1, \pm 2, \ldots)$ eine im weiteren Sinne stationäre Folge. Das Extrapolationsproblem läßt sich wie folgt formulieren: Man bestimme für ein ganzzahliges m eine lineare Funktion

$$\sum_{j=1}^{\infty} a_j X_{t-j},$$

die die Beziehung

$$E\left(\left| X_{t+m} - \sum_{j=1}^{\infty} a_j X_{t-j} \right|^2\right) = \min \tag{8.10.19}$$

erfüllt, und ermittle den Wert dieses Minimums. Mit anderen Worten, auf Grund der Kenntnis der Realisierung von X_t in den Zeitpunkten $t, t-1, t-2, \ldots$ wollen wir den Zustand dieses Prozesses zum Zeitpunkt $t + m$ so abschätzen, daß (8.10.19) erfüllt ist. Das ist im wesentlichen das Problem der Bestimmung der unbekannten Koeffizienten a_j nach der Methode der kleinsten Quadrate. Allerdings führt die unmittelbare Lösung nach der in 3.8 angegebenen Methode auf ein System von linearen Gleichungen mit unendlich vielen Unbekannten. Die Lösung eines solchen Systems ist aber bereits für eine endliche, jedoch große Anzahl von Unbekannten sehr mühsam. Kolmogoroff [12] und Wiener [3] haben ein anderes Lösungsverfahren angegeben, das auf der Beziehung zwischen der spektralen Verteilungsfunktion des stationären stochastischen Prozesses und der zugehörigen Korrelationsfunktion besteht.

Das Extrapolationsproblem für einen im weiteren Sinne stationären Prozeß mit kontinuierlicher Zeit läßt sich analog formulieren. Außerdem sind einige Arbeiten erschienen, in denen die hier behandelte Theorie der stationären stochastischen Prozesse auf mehrdimensionale stationäre Prozesse ausgedehnt wird. Eine systematische Darlegung der Begriffe und Ergebnisse auf diesem Gebiet findet man in den Arbeiten von Masani [1], Rosanow [1], Wiener und Masani [1]. Siehe auch Rosenblatt [2].

8.11. Martingale

Wir wollen nun kurz die Theorie eines anderen Typs stochastischer Prozesse streifen, die man als *Martingale* bezeichnet. Dieser Begriff wurde von Lévy [3] benutzt, jedoch verdanken wir das erste systematische Studium der Theorie und der Anwendungen der Martingale Doob [5], [7], [8], [9,] [11] (siehe auch Chow [1], Snell [1] und Billingsley [3]).

Definition 8.11.1. Eine Folge $\{X_n\}$ $(n = 1, 2, \ldots)$ von Zufallsvariablen nennt man ein *Martingal,* wenn für alle $n \geqq 1$ die folgenden Bedingungen erfüllt sind:

I. Es existiert der Erwartungswert $E(X_n)$ (und es ist somit $E(|X_n|) < \infty$).

II. Die Beziehung (vgl. 3.6.C)

$$E(X_{n+1} | x_1, \ldots, x_n) = X_n \tag{8.11.1}$$

gilt mit der Wahrscheinlichkeit 1.

Beispiel 8.11.1. Es sei $\{Y_j\}$ eine Folge unabhängiger Zufallsvariabler mit der Eigenschaft, daß $E(|Y_j|) < \infty$ $(j = 1, 2, \ldots)$ und $E(Y_j) = 0$ für $j > 1$ gilt. Wir setzen

$$X_n = \sum_{j=1}^{n} Y_j \quad (n = 1, 2, \ldots).$$

Die Folge $\{X_n\}$ ist ein Martingal. In der Tat, wegen der Unabhängigkeit der Zufallsvariablen Y_{n+1} und $(X_1, X_2, \ldots, X_n)$ erhalten wir

$$E(X_{n+1} | x_1, \ldots, x_n) = E[(X_n + Y_{n+1}) | x_1, \ldots, x_n] \tag{8.11.2}$$
$$= E(X_n | x_1, \ldots, x_n) + E(Y_{n+1}) = X_n.$$

Beispiel 8.11.2. Es möge x_1 (konstanter Wert) den Einsatz eines Spielers bedeuten, der sich an einem Glücksspiel beteiligt, und zwar zu Beginn der ersten Spielrunde. Ferner sei X_{n+1} $(n = 1, 2, \ldots)$ eine Zufallsvariable, die den Einsatz des Spielers nach der n-ten Spielrunde bezeichnet. Man pflegt zu sagen, daß es sich um ein gerechtes Spiel handelt, wenn der Erwartungswert für den Einsatz des Spielers nach jeder Spielrunde gleich dem Einsatz zu Beginn dieses Spieles ist. Mit anderen Worten (unter der Voraussetzung, daß $E(|X_n|) < \infty$ ist), ist das Spiel gerecht, wenn

$$E(X_{n+1} | x_1, \ldots, x_n) = X_n \quad (n = 1, 2, \ldots)$$

mit der Wahrscheinlichkeit 1 gilt. Man kann somit ein gerechtes Spiel als ein Martingal definieren.

Die Definition des Martingals für einen stochastischen Prozeß mit kontinuierlicher Zeit lautet wie folgt:

Definition 8.11.2. Einen stochastischen Prozeß $\{X_t, t \in I\}$ nennen wir ein *Martingal*, wenn $E(|X_t|) < \infty$ für alle $t \in I$ gilt und wenn für beliebige $t_1 < t_2 < \cdots < t_n$ $(t_j \in I;\ j = 1, 2, \ldots, n)$ die Beziehung

$$E(X_{t_{n+1}} | x_{t_1}, \ldots, x_{t_n}) = X_{t_n}$$

mit der Wahrscheinlichkeit 1 erfüllt ist.

Beispiel 8.11.3. Es sei $\{X_t, 0 \leqq t < \infty\}$ ein stochastischer Prozeß mit unabhängigen Zuwächsen, für den $P(X_0 = 0) = 1$ ist. Gilt für jedes $t > 0$ sowohl $E(|X_t|) < \infty$ als auch $E(X_t) = \text{const}$, so ist der Prozeß $\{X_t, 0 \leqq t < \infty\}$ für $t > 0$ ein Martingal. In der Tat, wir haben

$$X_{t_{n+1}} = X_{t_1} + (X_{t_2} - X_{t_1}) + \cdots + (X_{t_n} - X_{t_{n-1}}) + (X_{t_{n+1}} - X_{t_n}),$$

wobei die $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_{n+1}} - X_{t_n}$ unabhängig sind. Somit gilt

$$\begin{aligned} E(X_{t_{n+1}}|x_{t_1}, \ldots, x_{t_n}) &= E\{[X_{t_n} + (X_{t_{n+1}} - X_{t_n})]|x_{t_1}, \ldots, x_{t_n}\} \\ &= E(X_{t_n}|x_{t_1}, \ldots, x_{t_n}) + E(X_{t_{n+1}} - X_{t_n}) \\ &= X_{t_n} \end{aligned}$$

mit der Wahrscheinlichkeit 1.

8.12. Schlußbemerkungen

Dieses Kapitel war trotz seines großen Umfangs nur eine Einführung in die Theorie der stochastischen Prozesse. Ein Leser, der sich einen allgemeinen Überblick über die stochastischen Prozesse und ihre Anwendungsmöglichkeiten verschaffen will, wird mit Interesse die Fellersche Arbeit [6] lesen. Einem Leser, der sich gründlich mit der gesamten Theorie der stochastischen Prozesse und dem modernen mathematischen Rüstzeug dieser Theorie vertraut machen will, empfehlen wir die schöne Monographie von Doob [5]. Zusätzlich zu den zahlreichen Arbeiten über Markoffsche Prozesse, die wir früher erwähnt hatten, empfehlen wir insbesondere zwei Arbeiten von Hunt [1], [2] und eine Arbeit von Blumenthal [1]. Eine moderne Behandlung der Grundlagen und der ausgebauten Theorie der Markoffschen Prozesse findet man in den kürzlich erschienenen Monographien von Dynkin [3] und Chung [5]. Verallgemeinerungen von Markoffschen Prozessen, und zwar semi-Markoffsche Prozesse und Erneuerungsprozesse wurden von Lévy [10], Smith [1] und Pyke [2], [3] behandelt.

In dem Buch von Chintschin [2] und in den Arbeiten von Erdös und Kac [1], [2], LeCam [2], Donsker [1], [2], Prochorow [2], [3], Skorochod [1], Bartoszyński [1], [2] und Billingsley [1] sowie im zweiten Teil dieses Buches (vgl. 10.11.C) findet der Leser die Behandlung der Zusammenhänge zwischen den Grenzverteilungen für Summen von Zufallsvariablen und den Verteilungen, die in der Theorie der stochastischen Prozesse auftreten.

Wir erwähnen hier noch kurz einige Untersuchungen der Eigenschaften von Funktionen, welche Realisierungen eines stochastischen Prozesses darstellen, nämlich ihre Stetigkeit, eventuell der Charakter ihrer Unstetigkeitsstellen, ihre Integrierbarkeit usw.

Wiener [1] zeigte, daß der einzige Prozeß $\{X_t, 0 \leqq t < \infty\}$ mit unabhängigen und homogenen Zuwächsen, für den die Formeln (8.9.26) gelten und dessen Realisierungen mit der Wahrscheinlichkeit 1 stetige Funktionen sind, der Brownsche Bewegungsprozeß ist. Weitere Untersuchungen über Realisierungen eines Wienerschen Prozesses verdanken wir Lévy. Er stellte fest [8], daß diese Realisierungen irregulärer Natur sind, insbesondere, daß fast alle eine unbeschränkte Schwankung aufweisen. Eine weitere einschneidende Irregularität der Realisierungen eines Wienerschen Prozesses, und zwar die Tatsache, daß sie fast alle

an keiner Stelle zunehmen, haben kürzlich DVORETZKY, ERDÖS und KAKUTANI [1] nachgewiesen. Die Verteilungen von Funktionalen, die auf Realisierungen des Brownschen Bewegungsprozesses und verwandter Problemstellungen definiert sind, wurden von KAC [2], [4], DOOB [4], LÉVY [4], [6], FORTET [1] und anderen betrachtet. LÉVY [2] hat gezeigt, daß für jeden Prozeß X_t mit homogenen und unabhängigen Zuwächsen eine Funktion $f(t)$ derart existiert, daß fast alle Realisierungen des Prozesses $X_t - f(t)$ beschränkt sind und rechtsseitige sowie linksseitige Grenzwerte an jeder Stelle besitzen. Wir verweisen auf die besonders fruchtbaren Untersuchungen von DOEBLIN [1], [3], DOOB [2], [3], [5] und LÉVY [5], [7], die die Eigenschaften der Realisierungen Markoffscher Prozesse behandeln (siehe auch CHUNG [4]).

Es sei erwähnt, daß Bedingungen dafür, daß fast sämtliche Realisierungen eines Markoffschen Prozesses stückweise konstante Funktionen sind, von DOEBLIN [3] angegeben wurde; Bedingungen, unter denen sie stetig sind oder sowohl rechts- als auch linksseitige Ableitungen an jeder Stelle besitzen, gehen auf DYNKIN [1] und KINNEY [1] zurück (siehe auch SKOROCHOD [2]). URBANIK [1] untersuchte die Grenzwerteigenschaften der Realisierungen eines Markoffschen Prozesses für $t \to \infty$. Bedingungen, unter denen fast alle Realisierungen eines Markoffschen Prozesses mit endlich vielen Zuständen einen endlichen Erwartungswert für die Anzahl von Unstetigkeitsstellen haben, wurden von DOBRUSCHIN angegeben [2]. In der Monographie von DOOB [5] (Kap. XI, § 9) findet man eine Bedingung, unter der fast alle Realisierungen eines im weiteren Sinne stationären Prozesses absolut stetig sind. BELJAJEW [1] zeigte, daß fast alle Realisierungen eines Gaußschen stationären Prozesses mit einer stetigen Spektralfunktion im Intervall $(-\infty, \infty)$ unbeschränkt sind. Weitere Ergebnisse über die Realisierungen Gaußscher Prozesse findet man bei BELJAJEW [3], DOBRUSCHIN [6] und CIESIELSKI [1]. BELJAJEW [2] hat auch Bedingungen angegeben, unter denen fast alle Realisierungen eines im weiteren Sinne stationären Prozesses ganze Funktionen sind.

Wir erwähnen ferner eine Arbeit von MCKEAN [1] sowie zwei Arbeiten von BLUMENTHAL und GETOOR [1], [2], in denen Eigenschaften von Realisierungen einer gewissen Klasse stochastischer Prozesse mit unabhängigen Zuwächsen abgehandelt werden. Die Eigenschaften der Realisierungen von Martingalen wurden eingehend von DOOB [5], [7], [9] untersucht.

Für beliebige stochastische Prozesse, ohne die Voraussetzung, daß es sich um Markoffsche oder stationäre Prozesse handelt, notieren wir die folgenden Ergebnisse: KOLMOGOROFF (siehe SLUTSKI [2]) hat Bedingungen angegeben, unter denen fast alle Realisierungen eines stochastischen Prozesses stetig sind; TSCHENZOW [1] (siehe auch DYNKIN [1]) fand eine Bedingung, unter der fast alle Realisierungen linksseitige und rechtsseitige Grenzwerte besitzen; DOBRUSCHIN [5] und FISZ [7] gaben Bedingungen an, unter denen fast alle Realisierungen entweder stetig sind oder Unstetigkeitspunkte mit wenigstens einem einseitigen Grenzwert haben. FISZ [7] gab Bedingungen an, unter denen fast alle Reali-

sierungen stückweise konstante Funktionen mit einem endlichen Erwartungswert für die Anzahl der Unstetigkeitspunkte sind, wobei er gleichzeitig die Formel zur Berechnung des Erwartungswertes für die Anzahl der Unstetigkeitspunkte fand; schließlich haben Loève (siehe Lévy [4], mit einem von Loève verfaßten Anhang) und Beljajew [2] Bedingungen angegeben, unter denen fast alle Realisierungen stochastischer Prozesse, deren Zufallsvariable endliche Momente zweiter Ordnung haben, in der Umgebung eines gewissen Punktes analytisch sind. (Siehe die Aufgaben 8.13.21 bis 8.13.31.)

In den obigen Bemerkungen haben wir die gleichwertigen Redewendungen „fast alle" und „mit der Wahrscheinlichkeit 1" gebraucht, ohne zu präzisieren, was darunter zu verstehen ist, da dies die Besprechung der Konstruktion von Wahrscheinlichkeitsmaßen in reellen Funktionenräumen erfordert. Eine erschöpfende Behandlung dieser Fragen findet der Leser in den Arbeiten von Kolmogoroff [7] und Doob [1], [5]. Darüber hinaus sollte hinzugefügt werden, daß wir in den gegebenen Informationen die Voraussetzung weggelassen hatten, daß die betrachteten Prozesse separabel sind. Die präzise Definition dieses Begriffes findet man etwa bei Doob [5] (Kap. II, § 2); grob gesprochen bedeutet die Separabilität des Prozesses X_t, daß eine abzählbare Menge von Punkten $\{t_i\}$ $(i = 1, 2, \ldots)$ derart existiert, daß das Supremum oder Infimum fast aller Realisierungen von X_t in einem beliebigen offenen Intervall I gleich dem Supremum oder Infimum dieser Realisierung über der Teilmenge der Punkte t_i ist, die zum Intervall I gehören.

8.13. Aufgaben und Ergänzungen

1. Man zeige, daß der Satz 8.3.1 gültig bleibt, wenn man die Voraussetzung III durch die Beziehungen

$$\lim_{t\to 0} \frac{W_1(t)}{1 - W_0(t)} = 1 \quad \text{und} \quad \sum_{i=0}^{\infty} W_i(t) = 1 \quad \text{für jedes} \quad t > 0$$

ersetzt.

2. Man beweise, daß die Behauptung des Satzes 8.3.1 gültig bleibt, wenn man die Voraussetzung III durch die Voraussetzung ersetzt, daß die Anzahl der Signale in einem beliebigen endlichen Intervall $[0, t]$ endlich ist und daß in jedem Punkt höchstens ein Signal auftreten kann.

3. Es sei $\{X_t, 0 \leqq t < \infty\}$ ein Poissonscher Prozeß mit der Wahrscheinlichkeitsfunktion (8.3.4), und es bezeichne τ_k den (zufälligen) Punkt, in dem das k-te Signal aufgetreten ist.

a) Man beweise, daß für $j = 2, 3, \ldots$ sowie für beliebige $t_1, t_2, \ldots, t_j$ und T, für die $0 \leqq t_1 < t_2 < \cdots < t_j \leqq T$ gilt, die Beziehung

$$P(\tau_1 < t_1, \ldots, \tau_j < t_j \mid X_T = j) = \frac{j!}{T^j} \int_0^{t_1} \int_{t_1}^{t_2} \cdots \int_{t_{j-1}}^{t_j} d\tau_j \cdots d\tau_1$$

erfüllt ist (man vergleiche hierzu die Formel (10.7.1) mit $j_k = k$ und $f(x) = \frac{1}{T}$ für $0 \leqq x \leqq T$).

b) Man beweise, daß die Differenzen $U_k = \tau_k - \tau_{k-1}$ ($k = 1, 2, \ldots$) unabhängig sind und dieselbe durch

$$G(\tau) = P(U_k < \tau) = \begin{cases} 0 & \text{für } \tau \leqq 0, \\ 1 - \exp(-\lambda\tau) & \text{für } \tau > 0 \end{cases}$$

gegebene Exponentialverteilung haben.

Anmerkung. Jeden Signalprozeß $\{X_t, 0 \leqq t < \infty\}$ mit der Eigenschaft, daß die Differenzen U_k unabhängig sind und eine gemeinsame Verteilungsfunktion $G(\tau)$ haben, nennen wir einen *rekurrenten Prozeß*. Demnach ist der Poissonsche Prozeß ein Spezialfall eines rekurrenten Prozesses (siehe Takács [4], [6]).

4. Es sei $\{X_t, 0 \leqq t < \infty\}$ ein Signalprozeß, der die Voraussetzungen I und II des Satzes 8.3.1 erfüllt, und es seien seine Realisierungen rechtsseitig stetige Funktionen von t. Man zeige, daß die Beziehung (8.3.4) dann und nur dann gilt, wenn fast alle Realisierungen dieses Prozesses nur Sprünge der Sprunghöhe 1 aufweisen (Florek, Marczewski und Ryll-Nardzewski [1]).

5. Es sei $\{X_t, 0 \leqq t < \infty\}$ ein Signalprozeß, der die Voraussetzungen I und II des Satzes 8.3.1 erfüllt, und es sei $P(X_0 = 0) = 1$.

a) Man beweise: Gilt für ein gewisses ganzzahliges $l \geqq 1$ die Beziehung

$$\lim_{t \to 0} \frac{W_l(t)}{t} = \lambda_l \qquad (\lambda_l > 0)$$

und darüber hinaus

$$\lim_{t \to 0} \frac{1 - W_0(t) - W_l(t)}{t} = 0,$$

wobei die $W_l(t)$ durch (8.3.1) gegeben sind, so gilt für jedes $0 \leqq t < \infty$ und $i = 0, 1, 2, \ldots$

$$P(X_t = il) = \exp(-\lambda_l t)\, \frac{(\lambda_l t)^i}{i!}.$$

b) Man beweise: Gelten die Beziehungen

$$\lim_{t \to 0} \frac{W_1(t)}{t} = \lambda_1, \qquad \lim_{t \to 0} \frac{W_2(t)}{t} = \lambda_2,$$

$$\lim_{t \to 0} \frac{1 - W_0(t) - W_1(t) - W_2(t)}{t} = 0,$$

wobei $\lambda_1 > 0$ und $\lambda_2 > 0$ ist, so gilt für jedes $0 \leqq t < \infty$ und $i = 0, 1, 2, \ldots$

$$P(X_t = i) = \exp[-(\lambda_1 + \lambda_2)\, t] \sum_{r_1 + 2r_2 = i} \frac{(\lambda_1 t)^{r_1} (\lambda_2 t)^{r_2}}{r_1!\, r_2!}.$$

6. Es sei $\{X_t, 0 \leqq t < \infty\}$ ein Signalprozeß, der die Voraussetzungen I und II von 8.3 erfüllt, und es sei $P(X_0 = 0) = 1$. Ferner sei für jedes $t > 0$

$$\sum_{i=0}^{\infty} W_i(t) = 1,$$

wobei die $W_i(t)$ durch (8.3.1) gegeben sind. Man zeige, daß für jedes $0 \leqq t < \infty$ und $i = 0, 1, 2, \ldots$ die Formeln

$$W_i(t) = \exp\left(-t \sum_{k=1}^{\infty} \lambda_k\right) \sum_{\substack{r_1+2r_2+\cdots+lr_l=i \\ r_1\geqq 0,\ldots,r_l\geqq 0}} \frac{(\lambda_1 t)^{r_1} \cdots (\lambda_l t)^{r_l}}{r_1! \cdots r_l!}$$

gelten, wobei

$$\lim_{t\to 0} \frac{W_k(t)}{t} = \lambda_k \geqq 0 \qquad (k = 1, 2, \ldots),$$

$$\lim_{t\to 0} \frac{1 - W_0(t)}{t} = \sum_{k=1}^{\infty} \lambda_k < \infty$$

ist (Jánossy, Rényi und Aczél [1]). Man beachte, daß in dieser Aufgabe die Existenz der Intensitäten λ_k nicht vorausgesetzt wird.

7. a) Man beweise, daß sich der in Aufgabe 6 betrachtete Prozeß $\{X_t, 0 \leqq t < \infty\}$ in der Form

$$X_t = \sum_{k=1}^{\infty} k\, Y_{tk}$$

darstellen läßt, wobei $\{Y_{tk}, 0 \leqq t < \infty\}$ $(k = 1, 2, \ldots)$ ein Poissonscher Prozeß mit der Intensität λ_k ist; dabei sind für jedes $0 \leqq t < \infty$ die Zufallsvariablen $Y_{t1}, Y_{t2}, \ldots$ unabhängig.

b) Man zeige, daß die charakteristische Funktion $\varphi_t(u)$ der Zufallsvariablen X_t die folgende Form hat:

$$\varphi_t(u) = \exp\left[t \sum_{k=1}^{\infty} \lambda_k (e^{iuk} - 1)\right] = \sum_{k=1}^{\infty} \exp\left[\lambda_k t (e^{iuk} - 1)\right].$$

8. Es sei $\{X_k, 0 \leqq t < \infty\}$ der Signalprozeß von Aufgabe 5, und es möge die „Lebensdauer" des Signals eine Exponentialverteilung haben. Ferner sei $\{Y_t, 0 \leqq t < \infty\}$ die Anzahl der Signale, die zum Zeitpunkt t wirksam sind, und es sei $c_j(t) = P(Y_t = j)$. Man bestimme $c_j(t)$, $E(Y_t)$ und $D^2(Y_t)$ unter der Voraussetzung, daß $c_i(0) = 1$ ist.

9. Man beweise, daß die Erlangschen Formeln (8.5.17) gültig bleiben, wenn man die Voraussetzung, daß die Lebensdauer T der Signale eine Exponentialverteilung hat, durch die Annahme ersetzt, daß T eine beliebige Verteilung mit endlichem Erwartungswert besitzt (Fortet [2], Sewastjanow [2]).

10. Es sei $\{X_t, 0 \leqq t < \infty\}$ ein Poissonscher Prozeß mit der Intensität $\lambda > 0$. T_r sei die Zeit, die man auf das Auftreten des r-ten Signals $(r = 1, 2, \ldots)$ warten muß. Man zeige, daß T_r die durch die Formel (5.8.6) gegebene Gammaverteilung hat, wobei x durch t, b durch λ und p durch r ersetzt wird.

11. Es seien $\{q_i\}$ und $\{q_{ij}\}$ beliebige Folgen nichtnegativer Konstanten, und es sei

$$q_i = \sum_{j \neq i} q_{ij} \qquad (i = 1, 2, \ldots). \tag{*}$$

Man zeige, daß die q_i und die q_{ij} Intensitäten eines gewissen homogenen Markoffschen Prozesses sind (Doob [3]).

12. Wir betrachten einen homogenen Markoffschen Prozeß mit endlich vielen Zuständen mit der Übergangswahrscheinlichkeitsfunktion $p_{ij}(t)$ und endlichen Intensitäten q_i und q_{ij}, die der Gleichung (*) der vorhergehenden Aufgabe genügen. Wir bezeichnen mit ${}_np_{ij}(t)$ die Übergangswahrscheinlichkeit für den Übergang vom Zustand i nach dem Zustand j im Laufe der Zeit t, wenn in diesem Zeitpunkt n Zustandswechsel aufgetreten sind. Man zeige (Doob [2]), daß die folgenden Formeln gelten:

$${}_0p_{ij}(t) = \begin{cases} \exp(-q_i t) & \text{für } j = i, \\ 0 & \text{für } j \neq i, \end{cases}$$

$${}_{n+1}p_{ij}(t) = \sum_{k \neq i} \int_0^t \exp[-q_i(t-s)]\, q_{ik}\, {}_np_{kj}(s)\, ds.$$

13. Man zeige: Ersetzt man die Beziehung (8.9.4) durch die Beziehung

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|u| \geqq \varepsilon} f(t, t + \Delta t, u)\, du = 0,$$

so bleibt die Behauptung des Satzes 8.9.1 gültig.

14. Es sei $\{X_t, t \in I\}$, wobei I ein endliches Intervall bedeutet, ein Prozeß mit unabhängigen und homogenen Zuwächsen, und es sei $P(X_0 = 0) = 1$. Man zeige, daß dann für jedes $t \in I$ die Zufallsvariable X_t eine unbeschränkt teilbare Verteilung besitzt (siehe Aufgabe 5.14.25).

15. Es sei $\{U_n\}$ quadratisch im Mittel konvergent gegen die Zufallsvariable U. Man zeige, daß dann $\{U_n\}$ stochastisch gegen U konvergiert.

In den nachfolgenden Aufgaben 16 und 17 ist in den Beziehungen (*) stets die Konvergenz im quadratischen Mittel gemeint.

16. Es sei $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ eine im weiteren Sinne stationäre Folge, und es sei $m = 0$ und $\sigma = 1$; $R(\tau)$ sei die Korrelationsfunktion des Prozesses X_t. Man zeige, daß die Beziehung

$$\lim_{N \to \infty} \frac{1}{N+1} \sum_{t=0}^{N} X_t = 0 \tag{*}$$

dann und nur dann erfüllt ist (Slutski [2]), wenn

$$\lim_{N \to \infty} \frac{1}{N+1} \sum_{\tau=0}^{N} R(\tau) = 0$$

gilt.

17. Es sei $\{X_t, -\infty < t < \infty\}$ ein im weiteren Sinne stationärer Prozeß mit $m = 0$ und $\sigma = 1$. Wir setzen

$$\int_0^T X_t\,dt = \lim_{N\to\infty} \frac{T}{N} \sum_{k=1}^{N} X_{kT/N}. \tag{*}$$

Man zeige, daß die Beziehung

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T X_t\,dt = 0 \tag{*}$$

dann und nur dann gilt, wenn

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T R(\tau)\,d\tau = 0$$

erfüllt ist.

18. Es sei $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ eine im weiteren Sinne stationäre Folge mit der Korrelationsfunktion $R(\tau) = a^{|\tau|}$ $(\tau = 0, \pm 1, \pm 2, \ldots)$, wobei a reell und $|a| < 1$ ist. Man bestimme die Spektraldichte dieses Prozesses.

19. Die Spektraldichte der im weiteren Sinne stationären Folge $\{X_t, t = 0, \pm 1, \pm 2, \ldots\}$ habe die Form

$$f(\tau) = \frac{c}{2\pi} \frac{(e^{i\lambda} - b)(e^{-i\lambda} - b)}{(e^{i\lambda} - a)(e^{-i\lambda} - a)},$$

wobei a und b reell sowie $|a| < 1$ und $|b| < 1$ sind. Man zeige, daß die Korrelationsfunktion $R(\tau)$ die folgende Form hat:

$$R(\tau) = \begin{cases} c\,\dfrac{(1 - 2ab + b^2)}{1 - a^2} & \text{für } \tau = 0, \\[2ex] c\,\dfrac{(a - b)(1 - ab)}{1 - a^2}\, a^{|\tau| - 1} & \text{für } \tau = \pm 1, \pm 2, \ldots \end{cases}$$

20. Es sei $\{X_t, -\infty < t < \infty\}$ ein im weiteren Sinne stationärer Prozeß.

a) Man bestimme die Spektraldichte des Prozesses, wenn

$$R(\tau) = e^{-a|\tau|}$$

mit $a > 0$ gilt.

b) Man bestimme die Spektraldichte des Prozesses für

$$R(\tau) = e^{-a|\tau|} \cos b\tau \qquad (a > 0).$$

21. Es sei $\{X_t, 0 \leq t < \infty\}$ ein separabler Markoffscher Prozeß, und es sei die Beziehung

$$\lim_{\Delta t\to 0} P(X_{t+\Delta t} - X_t \neq 0 \mid X_t = x) = 0$$

gleichmäßig bezüglich t $(0 \leqq t < \infty)$ und bezüglich x $(-\infty < x < \infty)$ erfüllt. Man zeige, daß dann fast alle Realisierungen dieses Prozesses stückweise konstant mit endlich vielen Unstetigkeitspunkten in jedem endlichen Intervall sind (DOEBLIN [3]).

22. Es sei $\{X_t, t \in I\}$ ein separabler Markoffscher Prozeß, wobei I ein abgeschlossenes endliches Intervall ist. Wir setzen

$$C(h, \varepsilon) = \sup_{\substack{t, t+\Delta t \in I \\ -\infty < x < \infty}} P(|X_{t+\Delta t} - X_t| > \varepsilon \,|\, X_t = x).$$

a) Man beweise: Gilt für jedes $\varepsilon > 0$ die Beziehung

$$\lim_{h \to 0} \frac{C(h, \varepsilon)}{h} = 0,$$

so sind fast alle Realisierungen des Prozesses stetig in I. (DYNKIN [1], KINNEY [1]. Diese Behauptung gilt auch unter schwächeren Voraussetzungen. Siehe FISZ [7].)

b) Man zeige: Gilt für jedes $\varepsilon > 0$ die Beziehung

$$\lim_{h \to 0} C(h, \varepsilon) = 0,$$

so haben fast alle Realisierungen des Prozesses beide einseitigen Grenzwerte an jeder Stelle $t \in I$ (KINNEY [1]).

23. a) Es sei $\{X_t, 0 \leqq t \leqq 1\}$ ein stochastischer separabler Prozeß. Man beweise: Gilt die Beziehung

$$E(|X_{t+\alpha} - X_t|^\beta) < C|\Delta t|^{1+\alpha}, \qquad (*)$$

wobei $\beta > 0, \alpha > 0$ und C Konstanten sind, für beliebige Punkte t und $t + \Delta t$ aus dem Intervall (0, 1], so ist für jedes $\gamma < \frac{\alpha}{\beta}$ die Beziehung

$$P\left(\lim_{\Delta t \to 0} \frac{X_{t+\Delta t} - X_t}{|\Delta t|^\gamma} = 0, \quad 0 \leqq t \leqq 1\right) = 1 \qquad (**)$$

gleichmäßig bezüglich t erfüllt. (Die Stetigkeit fast aller Realisierungen des Prozesses X_t unter der Voraussetzung (*) hat KOLMOGOROFF nachgewiesen. Den Beweis dafür findet man z. B. in der Arbeit von SLUTSKI [2]. Ein Satz mit einer der Behauptung (**) äquivalenten Aussage stammt von LOÈVE [4], 3. Auflage.)

b) Aus der Beziehung (**) leite man her, daß für $\beta < \alpha$

$$P(X_t = \text{const}, 0 \leqq t \leqq 1) = 1$$

gilt.

c) Man beweise, daß für den Wienerschen Prozeß die Beziehung (**) für jedes $\gamma < \frac{1}{2}$ gilt.

24. Es sei $\{X_t, 0 \leqq t \leqq 1\}$ ein separabler stochastischer Prozeß. Wir wollen annehmen, daß für beliebige Punkte t_1, t_2, t_3 $(t_1 < t_2 < t_3)$ aus dem Intervall [0, 1] die Beziehung

$$E[|X_{t_1} - X_{t_2}|^p |X_{t_2} - X_{t_3}|^q] < C|t_1 - t_3|^{1+r}$$

gilt, wobei $p > 0$, $q > 0$ und $r > 0$ und C von t unabhängig ist. Man zeige, daß dann fast alle Realisierungen des Prozesses beide einseitige Ableitungen an jeder Stelle $t \in [0, 1]$ besitzen (Tschenzow [1]).

25. Es sei $\{X_t, t \in I_0\}$ ein separabler stochastischer Prozeß, wobei I_0 ein endliches abgeschlossenes Intervall bezeichnet. X_I möge den Zuwachs des Prozesses X_t im Intervall $I \subset I_0$ bezeichnen. Ferner sei $a(I) = P(X_I \neq 0)$, $b(I, \varepsilon) = P(|X_I| > \varepsilon)$, während $|I|$ die Länge des Intervalls I bedeuten möge. Wir nehmen an, daß die Beziehung

$$\lim_{|I| \to 0} b(I, \varepsilon) = 0 \tag{*}$$

für jedes $\varepsilon > 0$ gilt und daß darüber hinaus

$$\overline{\lim_{n \to \infty}} \sum_{k=1}^{n} a(I_{nk}) < \infty \tag{**}$$

ist, wenn $\max\limits_{1 \leq k \leq n} |I_{nk}| \to 0$ geht; hierbei bedeutet $\{I_{nk}\}$ eine Unterteilung des Intervalls I_0 in disjunkte Teilintervalle I_{nk} $(k = 1, 2, \ldots, n)$. Man zeige, daß dann fast alle Realisierungen des Prozesses in I_0 stückweise konstant sind und daß die Anzahl der Unstetigkeitspunkte einen endlichen Erwartungswert besitzt, der dem Grenzwert auf der linken Seite von (**) gleich ist. Darüber hinaus sind in jedem einzelnen Punkt $t \in I_0$ die Realisierungen mit der Wahrscheinlichkeit 1 stetig (Fisz [3], [7]).

26. Es sei $\{X_t, t \in I_0\}$ ein separabler stochastischer Prozeß. Man zeige, daß fast alle Realisierungen des Prozesses keine Unstetigkeiten erster Art[1]) besitzen, wenn für jedes $\varepsilon > 0$ die Beziehung

$$\overline{\lim_{n \to \infty}} \sum_{k=1}^{n} b(I_{nk}, \varepsilon) = 0 \tag{*}$$

erfüllt ist, sobald $\max\limits_{1 \leq k \leq n} |I_{nk}| \to 0$ geht; hierbei hat $\{I_{nk}\}$ die gleiche Bedeutung wie in Aufgabe 25 (Fisz [7]).

27. Es sei $\{X_t, t \in I_0\}$ ein separabler homogener Markoffscher Prozeß oder ein separables Martingal. Man zeige, daß bei Gültigkeit der Beziehung (*) von Aufgabe 26 fast alle Realisierungen des Prozesses stetig sind (wegen des Markoffschen Prozesses vergleiche Fuchs [1], wegen der Martingale Fisz [7]).

28. Es sei $\{X_t, t \in I_0\}$ ein separabler Prozeß mit unabhängigen Zuwächsen, und es sei

$$\lim_{|I| \to 0} a(I) = 0.$$

Man zeige, daß dann für fast alle Realisierungen des Prozesses sämtliche Behauptungen der Aufgabe 25 gelten.

29. Man zeige, daß fast alle Realisierungen eines reellen, separablen, stationären Gaußschen Prozesses $\{X_t, -\infty < t < \infty\}$ mit stetiger spektraler Verteilungsfunktion unbeschränkt sind (Beljajew [1]).

---

[1]) Die Funktion $f(t)$ hat eine Unstetigkeit erster Art an der Stelle t, wenn beide Grenzwerte $f(t - 0)$ und $f(t + 0)$ existieren und voneinander verschieden sind.

30. Es sei $\{X_t, -\infty < t < \infty\}$ ein separabler, reeller, stationärer und stetiger Gaußscher Prozeß. Man zeige (DOBRUSCHIN [6]), daß dann entweder fast alle Realisierungen stetig sind oder aber $\beta > 0$ derart existieren, daß fast alle Realisierungen für jedes t_0 die Beziehung

$$\overline{\lim_{t\to t_0}}\, X_t - \underline{\lim_{t\to t_0}}\, X_t \geqq \beta$$

erfüllen.

31. Es sei $\{X_t, -\infty < t < \infty\}$ ein reeller, separabler und stetiger, im weiteren Sinne stationärer Prozeß mit $E(X_t) = 0$ und $D^2(X_t) = 0$. Man zeige: Genügt die Korrelationsfunktion $R(\tau)$ für $\tau \to 0$ der Beziehung[1])

$$R(\tau) = 1 - O(|\tau|^{1+\delta})$$

mit $\delta > 0$, so sind fast alle Realisierungen des Prozesses stetig.

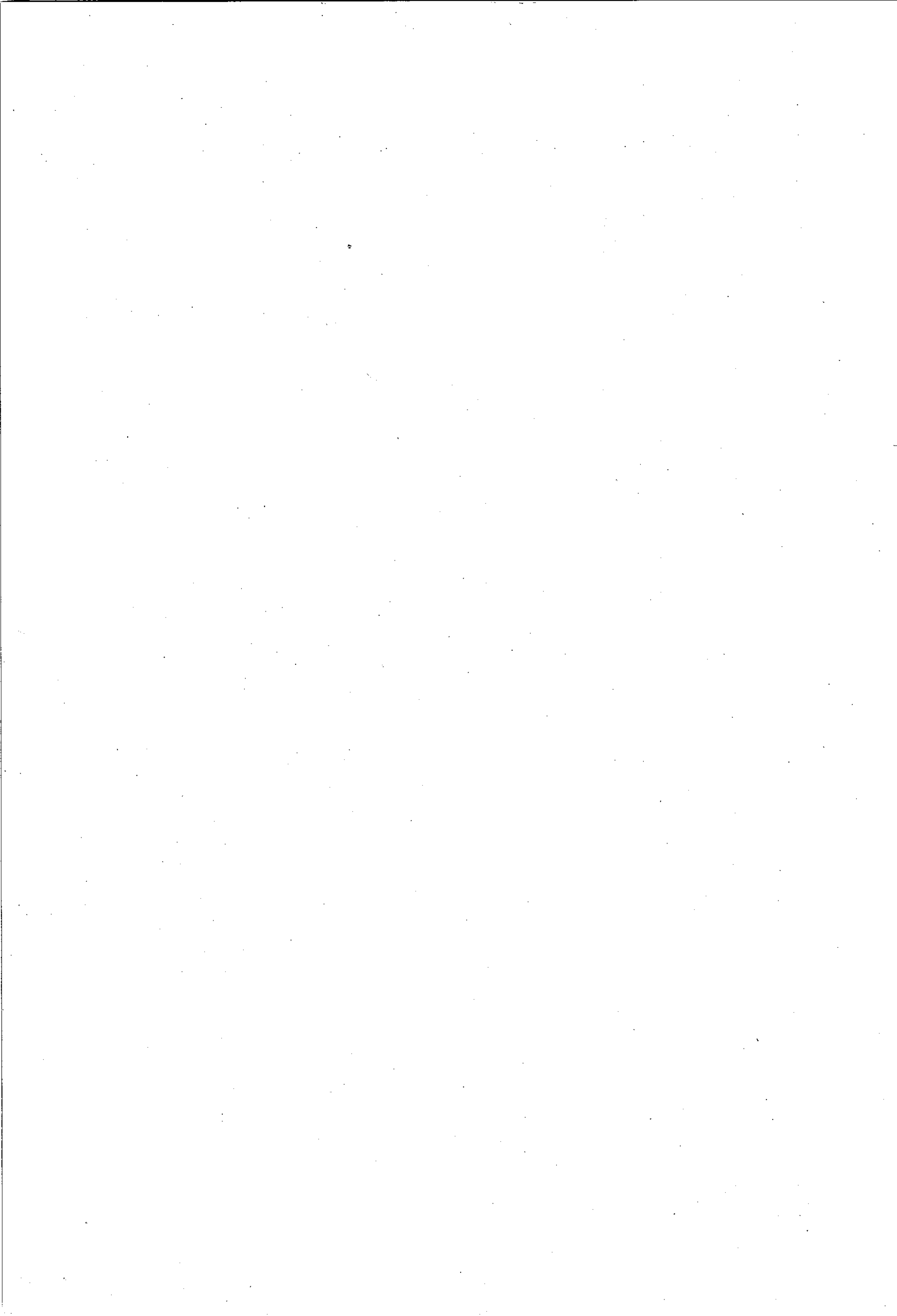
Hinweis. Man stütze sich auf die Aufgabe 23.

---

[1]) Mit anderen Worten, es existiert eine Konstante $C > 0$ derart, daß für $\tau \to 0$ die Beziehung $R(\tau) < C|\tau|^{1+\delta}$ gilt.

ZWEITER TEIL

MATHEMATISCHE STATISTIK



9. STICHPROBENMOMENTE UND IHRE FUNKTIONEN

9.1. Der Begriff einer Stichprobe

Im ersten Teil dieses Buches wurde die Wahrscheinlichkeitsrechnung behandelt. Wir bauten die Wahrscheinlichkeitsrechnung als eine mathematische Theorie auf, in welcher der abstrakte Wahrscheinlichkeitsbegriff dem Häufigkeitsbegriff zufälliger Erscheinungen entspricht, die man in der Wirklichkeit beobachten kann. Der Wahrscheinlichkeitsbegriff steht in demselben Verhältnis zur Häufigkeit eines zufälligen Ereignisses wie etwa der exakte Begriff einer geometrischen Figur in der euklidischen Geometrie zur wirklichen Figur im uns umgebenden Raum. Deshalb hat die Wahrscheinlichkeitsrechnung eine ähnliche Bedeutung für statistische Probleme, bei welchen man durch Beobachtung von Häufigkeiten die Gesetzmäßigkeiten zufälliger Erscheinungen untersucht, wie sie die Geometrie für die Geodäsie hat.

Der zweite Teil dieses Buches will zeigen, wie mit Hilfe der Wahrscheinlichkeitsrechnung zahlreiche statistische Probleme gelöst werden können.

Im Laufe dieser Untersuchungen wird es notwendig sein, unsere Kenntnisse über die Wahrscheinlichkeitsrechnung zu vertiefen, vor allem auf den Gebieten, in denen sich Wahrscheinlichkeitsrechnung und mathematische Statistik eng berühren.

Die statistischen Fragen, mit denen wir uns befassen wollen, bestehen meistens darin, daß auf Grund der Kenntnisse der Eigenschaften einer entsprechend gewählten Teilmenge von Elementen, die einer gewissen Gesamtmenge entnommen wurden, Schlüsse zu ziehen sind, welche die zu untersuchenden Eigenschaften der übrigen unbekannten Elemente betreffen.

In der Statistik nennt man die Gesamtmenge, deren Elemente wir jeweils untersuchen, die *Grundgesamtheit* oder einfach die *Gesamtheit.*

Die Elemente der Gesamtheit kann man hinsichtlich verschiedener Merkmale untersuchen. Wenn wir sagen, daß die Grundgesamtheit die Verteilung $F(x)$ hat, so wollen wir dadurch ausdrücken, daß wir das Merkmal X der Elemente dieser Grundgesamtheit untersuchen und dieses Merkmal in der Gesamtheit eine Verteilung hat, die durch die Verteilungsfunktion $F(x)$ charakterisiert wird.

In der Statistik nennt man eine endliche Teilmenge, die aus Elementen der Grundgesamtheit besteht, *Stichprobe.* Wir wollen unter einer Stichprobe eine Folge von Merkmalswerten einer gewissen Anzahl von Elementen, die zur Gesamtheit gehören, verstehen. Die Elementeanzahl der Stichprobe wollen wir ihren *Umfang* nennen. Wir fragen nun nach der mathematischen Bedeutung einer *zufälligen Stichprobe.*

Es sei das Merkmal X der Elemente einer Grundgesamtheit eine Zufallsvariable mit der Verteilungsfunktion $F_0(x)$. Wir betrachten eine Gesamtheit, deren Elemente j-elementige Systeme ($j = 2, 3, \ldots$) sind, die aus allen möglichen Elementen der Grundgesamtheit gebildet werden. Die Elemente dieser m-dimensionalen Gesamtheit charakterisiert der Zufallsvektor $(X_1', X_2', \ldots, X_j')$, wobei X_k' ($k = 1, 2, \ldots, j$) der Wert des Merkmals X des k-ten Elements des Systems ist. Wir wollen annehmen, daß für $k = 1, 2, \ldots$ sowie für beliebige $x_1, x_2, \ldots, x_k$ die bedingte Wahrscheinlichkeit

$$F_k(x \mid x_1, \ldots, x_k) = P(X_{k+1}' < x \mid X_1' = x_1, \ldots, X_k' = x_k)$$

existiert. Nach einer gewissen Methode, wir wollen sie mit M bezeichnen, wählen wir aus der Grundgesamtheit Systeme von je n Elementen als Stichproben aus und beobachten die Werte $x_1, x_2, \ldots, x_n$ des Merkmals X der ausgewählten Elemente.

Es sei jetzt $(X_1, X_2, \ldots, X_n)$ der beobachtete Zufallsvektor, d. h., X_k ($k = 1, 2, \ldots, n$) ist eine Zufallsvariable, deren mögliche Wertmenge aus den Beobachtungen an x_k, dem k-ten Stichprobenelement, in jeder möglichen n-elementigen Stichprobe besteht, die nach unserer Methode M aus der Grundgesamtheit ausgewählt wurde.

Definition 9.1.1. Die Auswahlmethode M von Elementen in die Stichprobe ist *zufällig*, wenn

1. für jedes x die Gleichung

$$P(X_1 < x) = F_0(x) \tag{9.1.1}$$

gilt;

2. für jedes $k = 1, 2, \ldots, n - 1$ und für beliebige $x_1, x_2, \ldots, x_k$ die Gleichung

$$P(X_{k+1} < x \mid X_1 = x_1, \ldots, X_k = x_k) = F_k(x \mid x_1, \ldots, x_k) \tag{9.1.2}$$

erfüllt ist.

Eine Stichprobe $(x_1, x_2, \ldots, x_n)$, die wir nach einer zufälligen Auswahlmethode erhalten haben, nennen wir eine *zufällige Stichprobe.*

Dies läßt sich geometrisch deuten: Alle möglichen Werte der n-dimensionalen Zufallsvariablen $(X_1, X_2, \ldots, X_n)$ stellen eine Punktmenge im n-dimensionalen Raum dar. Eine zufällige Stichprobe vom Umfang n ist ein Punkt dieser Punktmenge. Die Menge aller zufälligen Stichproben vom Umfang n nennen wir den *Stichprobenraum.*

Methoden der Stichprobenerhebung werden in Kapitel 14 behandelt.

Definition 9.1.2. Eine zufällige Stichprobe nennen wir *einfach,* wenn die Zufallsvariablen $X_1, X_2, \ldots, X_n$ voneinander unabhängig sind.

Sind andere Voraussetzungen nicht besonders hervorgehoben, so verstehen wir unter einer *Stichprobe* immer eine zufällige Stichprobe.

Beispiel 9.1.1. Für jede durch eine staatliche Landwirtschaftszählung erfaßte Bauernwirtschaft wurde eine Karteikarte angefertigt. Auf diesen Karten war unter den untersuchten

Merkmalen auch die Größe des zu einer Bauernwirtschaft gehörenden Ackerlandes in Hektar angegeben. Alle Bauernwirtschaften hatte man hinsichtlich dieses Merkmals in 7 Gruppen eingeteilt, die durch die Zahlen 1 bis 7 gekennzeichnet wurden. Auf jeder Karteikarte ist also eine der Zahlen 1 bis 7 eingetragen, die wir mit X bezeichnen wollen. Welchen Wert X annimmt, hängt davon ab, zu welcher Gruppe der untersuchte landwirtschaftliche Betrieb gehört. Die Größe X sei das zu untersuchende Merkmal.

Die Grundgesamtheit wird hier durch alle Karteikarten dargestellt; ihre Anzahl sei N. Aus dieser Kartenmenge greifen wir aufs Geratewohl $n = 10$ Karten heraus; dabei legen wir aber vor Wahl jeder nächsten Karte die vorher herausgegriffene Karte in die Gesamtmenge der Karten zurück. Die Wahrscheinlichkeit dafür, daß das Merkmal X den Wert i ($i = 1, \ldots, 7$) annimmt, ändert sich also während des Herausgreifens der einzelnen Karten nicht. Die Werte $(x_1, x_2, \ldots, x_{10})$ der Zufallsvariablen X, die auf den 10 herausgegriffenen Karteikarten beobachtet wurden, stellen einen Wert einer zehndimensionalen Zufallsvariablen dar. Dieser beobachtete Wert der zehndimensionalen Zufallsvariablen ist eine Stichprobe vom Umfang 10. Der Stichprobenraum besteht aus allen möglichen 10-Tupeln $(x_1, x_2, \ldots, x_{10})$. Da x_k ($k = 1, 2, \ldots, 10$) eine der Zahlen $1, 2, \ldots, 7$ darstellt, besteht hier der Stichprobenraum aus 7^{10} Punkten.

9.2. Der Begriff der Stichprobenfunktion

Definition 9.2.1. Unter einer *Stichprobenfunktion* verstehen wir eine Zufallsvariable, die selbst eine Funktion einer Stichprobe, d. h. einer beobachteten n-dimensionalen Zufallsvariablen $(X_1, X_2, \ldots, X_n)$ ist.

Die Frage nach der Verteilung einer Stichprobenfunktion gehört zu den Grundproblemen der mathematischen Statistik.

Hier interessieren uns zwei derartige Probleme. Das eine lautet: Wir suchen für jedes natürliche n die Verteilung der gegebenen Stichprobenfunktionen $Z_n = Z(X_1, X_2, \ldots, X_n)$; wir wollen sie die *exakte Verteilung der Stichprobenfunktion* nennen. Die Kenntnis der exakten Verteilung einer Stichprobenfunktion hat große praktische Bedeutung für statistische Untersuchungen, in denen die Anzahl der Beobachtungen klein ist. Wir sprechen in diesem Fall von einer *kleinen Stichprobe.*

Bei der zweiten Fragestellung suchen wir nicht die Verteilungen einer Stichprobenfunktion für jedes einzelne n zu bestimmen, sondern interessieren uns nur für die Grenzverteilung der Stichprobenfunktion Z_n für $n \to \infty$. Die Grenzverteilungen der Stichprobenfunktionen finden dort Anwendung, wo die Anzahl der Beobachtungen groß ist. Wir sprechen dann von einer *großen Stichprobe.*

Es existiert kein allgemeines Kriterium, das uns zu entscheiden erlaubt, wann eine Stichprobe als groß und wann sie als klein anzusehen ist. Diese Entscheidung hängt von den einzelnen Stichprobenfunktionen ab. Eine Stichprobe kann für die eine Stichprobenfunktion als groß gelten, während sie für eine andere Stichprobenfunktion noch nicht als genügend groß angesehen werden kann.

9.3. Die Verteilung des arithmetischen Mittels normal verteilter Zufallsvariabler

Die Bestimmung der exakten Verteilungen einer Stichprobenfunktion ist im allgemeinen sehr kompliziert. Doch gibt es verhältnismäßig einfache Methoden, um dieses Problem in einigen wichtigen Sonderfällen zu lösen.

Vor allem wollen wir die Verteilung einiger Stichprobenfunktionen bestimmen, falls das betrachtete Merkmal X der Grundgesamtheit normal verteilt ist mit der Dichte

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right\}; \tag{9.3.1}$$

hierbei ist m der Mittelwert und σ die Standardabweichung von X. Wir betrachten die durch

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{9.3.2}$$

definierte Stichprobenfunktion und setzen voraus, daß die Zufallsvariablen X_k unabhängig sind und alle dieselbe durch die Dichte (9.3.1) bestimmte Verteilung besitzen. Um nun die Verteilung dieser Stichprobenfunktion zu finden, wenden wir die Theorie der charakteristischen Funktionen an. Wie aus (5.7.4) bekannt ist, hat die charakteristische Funktion der normalen Zufallsvariablen X die Form

$$\varphi(t) = \exp\left(-\frac{t^2\sigma^2}{2} + itm\right). \tag{9.3.3}$$

Die Zufallsvariablen X_k sind unabhängig, und daher folgt aus den Gleichungen (4.4.3) und (4.2.15), daß die charakteristische Funktion $\varphi_1(t)$ der Stichprobenfunktion $\overline{X}$ die Form

$$\varphi_1(t) = \exp\left(-\frac{t^2\sigma^2}{2n} + itm\right) \tag{9.3.4}$$

hat. Dies ist die charakteristische Funktion einer normalen Zufallsvariablen $N\left(m; \frac{\sigma}{\sqrt{n}}\right)$. Also wird die Dichte der Zufallsvariablen $\overline{X}$ durch die Formel

$$f_1(\overline{x}) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n}{2}\frac{(\overline{x}-m)^2}{\sigma^2}\right) \tag{9.3.5}$$

ausgedrückt. Wie wir sehen, hat die Zufallsvariable $\overline{X}$ denselben Mittelwert wie X; sie ist aber mehr um diesen Mittelwert konzentriert.

Beispiel 9.3.1. Die Zufallsvariablen X_k sind unabhängig und haben alle die gleiche Dichte, die durch folgende Formel bestimmt ist:

$$f(x) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-1)^2}{4}\right). \tag{9.3.6}$$

Wir suchen die Verteilung der Zufallsvariablen

$$\overline{X} = \frac{1}{16}\sum_{k=1}^{16} X_k.$$

Nach (9.3.5) hat die Zufallsvariable $\overline{X}$ die Dichte

$$f_1(\overline{x}) = \frac{2}{\sqrt{2\pi}} \exp\left(-2(\overline{x}-1)^2\right).$$

Die Standardabweichung der Zufallsvariablen $\overline{X}$ beträgt also 0,5. Wir suchen die Wahrscheinlichkeit dafür, daß $0 \leqq \overline{X} \leqq 2$ ist. Es gilt

$$P(0 \leqq \overline{X} \leqq 2) = P\left(-2 \leqq \frac{\overline{X}-1}{0{,}5} \leqq 2\right).$$

Den Tafeln für die Normalverteilung entnehmen wir

$$P(0 \leqq \overline{X} \leqq 2) \approx 0{,}9544.$$

Zum Vergleich berechnen wir die Wahrscheinlichkeit dafür, daß die Zufallsvariable X in denselben Grenzen enthalten ist. Wir erhalten

$$P(0 \leqq X \leqq 2) = P\left(-\frac{1}{2} \leqq \frac{X-1}{2} \leqq \frac{1}{2}\right) \approx 0{,}3830.$$

Wie wir sehen, ist die Zufallsvariable $\overline{X}$ weniger gestreut als X.

Dieses Beispiel kann man praktisch so deuten: Es liegt eine bestimmte Gesamtheit vor, und uns interessiert das Merkmal X der Elemente dieser Gesamtheit.

Dieses Merkmal habe eine durch (9.3.6) bestimmte Normalverteilung. Aus der Gesamtheit greifen wir aufs Geratewohl unabhängige einfache Stichproben zu 16 Elementen heraus, beobachten die Werte des Merkmals $\overline{X}$ in diesen Stichproben und berechnen den Wert der Stichprobenfunktion X. Es zeigte sich, daß wir bei genügend oft wiederholter Entnahme dieser Stichprobe vom Umfang 16 in ungefähr 95 von 100 Fällen solche Werte von $\overline{X}$ erhalten, die nicht negativ und nicht größer als 2 sind. Wählten wir dagegen wiederholt ein einziges Element, dann erhielten wir nur in ungefähr 38 von 100 Fällen Werte von X, die in denselben Grenzen liegen.

Der Formel (9.3.5) entnehmen wir, daß die Verteilung der Stichprobenfunktion $\overline{X}$ von n abhängt. In Abb. 5.7.1 sind die Verteilungen $N(0; 1)$, $N(0; 0{,}5)$ und $N(0; 0{,}25)$ graphisch dargestellt. Sie stellen gleichzeitig die durch (9.3.5) bestimmten Dichten von X für $n = 1$, $n = 4$, $n = 16$ dar, wenn die Veränderlichen X_k nach der Formel (9.3.1) verteilt sind, in der $m = 0$ und $\sigma = 1$ gesetzt ist.

9.4. Die χ^2-Verteilung

Wir betrachten n unabhängige Zufallsvariable X_k mit derselben Verteilung, deren Dichte

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) \tag{9.4.1}$$

sei. Die Voraussetzung $E(X_k) = 0$ schränkt hier die Allgemeinheit nicht ein; wäre nämlich $E(X_k) = m \neq 0$, so könnten wir die Zufallsvariablen $X_k - m$ betrachten, deren Mittelwert $E(X_k - m)$ gleich Null ist.

Die Stichprobenfunktion χ^2 definieren wir durch

$$\chi^2 = \sum_{k=1}^{n} X_k^2. \tag{9.4.2}$$

Beispiel 9.4.1. Das Merkmal X der Elemente einer gewissen Gesamtheit sei normal verteilt mit dem Mittelwert $E(X) = 0$. Dieser Gesamtheit entnehmen wir eine Stichprobe vom Umfang n und beobachten an deren Elementen die Werte des Merkmals X. Wir bezeichnen diese Werte mit $x_1, x_2, \ldots, x_n$ und berechnen den Wert des Ausdrucks $\sum_{k=1}^{n} x_k^2$. Dies ist der beobachtete Wert der Stichprobenfunktion χ^2.

Wir suchen zuerst die Verteilung der Zufallsvariablen $Y = X^2$, wenn die Verteilung von X durch (9.4.1) bestimmt ist. Auf Grund von (2.4.14) hat die Dichte der Zufallsvariablen Y die Form

$$g_1(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left[f(\sqrt{y}) + f(-\sqrt{y})\right] = \dfrac{1}{\sigma\sqrt{2\pi}}\, y^{-\frac{1}{2}}\, e^{-\frac{1}{2}\frac{y}{\sigma^2}} & \text{für} \quad y > 0, \\ 0 & \text{für} \quad y \leqq 0. \end{cases} \tag{9.4.3}$$

Die Zufallsvariable Y hat also eine Gammaverteilung nach der Formel (5.8.6), in der $p = \frac{1}{2}$, $b = \frac{1}{2\sigma^2}$ zu setzen ist. Da für Zufallsvariable mit einer Gammaverteilung der Additionssatz gilt (vgl. 5.8), hat die Zufallsvariable χ^2, die durch die Formel (9.4.2) gegeben ist, ebenfalls eine Gammaverteilung mit $p = \frac{n}{2}$ und $b = \frac{1}{2\sigma^2}$. Also ist ihre Dichte durch

$$g_n(u) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\sigma^n \Gamma\left(\frac{n}{2}\right)}\, e^{-\frac{u}{2\sigma^2}}\, u^{\frac{n}{2}-1} & \text{für} \quad u > 0, \\ 0 & \text{für} \quad u \leqq 0 \end{cases} \tag{9.4.4}$$

gegeben, wobei u die Werte der Zufallsvariablen χ^2 sind.

Die Verteilung der Zufallsvariablen wurde von Helmert [1] berechnet.

Aus (5.8.10) erhalten wir für den Mittelwert und die Dispersion der Zufallsvariablen χ^2

$$m_1 = n\sigma^2, \quad \mu_2 = 2n\sigma^4. \tag{9.4.5}$$

Die durch (9.4.4) bestimmte Verteilung nennt man die *χ^2-Verteilung mit n Freiheitsgraden*, entsprechend der Tatsache, daß χ^2 eine Summe von n unabhängigen Summanden ist.

Für zwei beliebige Werte $u_1 \geqq 0$ und $u_2 > 0$ mit $u_1 < u_2$ gilt also

$$P(u_1 \leqq \chi^2 < u_2) = \int_{u_1}^{u_2} g_n(u)\,du;$$

dabei ist $g_n(u)$ die durch (9.4.4) bestimmte Funktion.

In den Anwendungen der Stichprobenfunktion χ^2, die wir später besonders besprechen werden, benutzt man häufig den Ausdruck

$$P(\chi^2 \geqq u_0) = \frac{1}{2^{\frac{n}{2}}\sigma^n \Gamma\left(\frac{n}{2}\right)} \int_{u_0}^{\infty} e^{-\frac{u}{2\sigma^2}} u^{\frac{n}{2}-1}\,du, \tag{9.4.6}$$

in dem $u_0 > 0$ ist.

Die Tafeln der χ^2-Verteilung enthalten die Werte (9.4.6) für verschiedene Zahlen u_0 und n bei festem $\sigma = 1$. Um diese Tafeln für $\sigma \neq 1$ benutzen zu können, hat man $u_0 = z_0\sigma^2$ zu setzen. Dann geht $P(\chi^2 \geqq u_0)$ in $P(\chi^2 \geqq z_0\sigma^2)$ über. Führt man nämlich in dem Ausdruck (9.4.4) die Substitution $z = \dfrac{u}{\sigma^2}$ aus und beachtet man die Regel für die Transformation einer Dichte, so ergibt sich

$$g_n(z) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{z}{2}} z^{\frac{n}{2}-1} & \text{für} \quad z > 0, \\ 0 & \text{für} \quad z \leqq 0. \end{cases} \tag{9.4.7}$$

Der Ausdruck (9.4.7) ist schon die Dichte der Stichprobenfunktion χ^2 für $\sigma = 1$.

In Abb. 9.4.1 sind die Dichten der Zufallsvariablen χ^2 für einen Freiheitsgrad und für sechs Freiheitsgrade dargestellt.

Beispiel 9.4.2. Die Zufallsvariablen X_k $(k = 1, 2, \ldots, 8)$ seien unabhängig, und jede von ihnen habe die gleiche Normalverteilung $N(0; 2)$. Wir bestimmen die Stichprobenfunktion

$$\chi^2 = \sum_{k=1}^{8} X_k^2.$$

Die Zufallsvariable χ^2 hat also 8 Freiheitsgrade; ihr Mittelwert und ihre Standardabweichung sind nach (9.4.5)

$$m_1 = 32, \qquad \sqrt{\mu_2} = 16.$$

Wir berechnen die Wahrscheinlichkeit dafür, daß χ^2 nicht kleiner als 40 ist. Wegen $\sigma = 2$ ist $\frac{40}{\sigma^2} = 10$. Diese Wahrscheinlichkeit ist also gleich $P(\chi^2 \geqq 10)$. Aus den Tafeln der χ^2-Verteilung (für $\sigma = 1$) erhalten wir bei 8 Freiheitsgraden

$$P(\chi^2 \geqq 9{,}524) \approx 0{,}30, \qquad P(\chi^2 \geqq 11{,}030) \approx 0{,}20.$$

Bei linearer Interpolation ergibt sich

$$P(\chi^2 \geqq 10) \approx 0{,}30 - \frac{10 - 9{,}524}{11{,}030 - 9{,}524} \cdot 0{,}10 \approx 0{,}27.$$

Abb. 9.4.1

Beispiel 9.4.3. In Beispiel 5.11.1 stellten wir fest, daß die Geschwindigkeitskomponenten V_1, V_2, V_3 eines Gasmoleküls (in einem idealen Gas) die gemeinsame Dichte

$$h(v_1, v_2, v_3) = \frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} \exp\left(-\frac{v_1^2 + v_2^2 + v_3^2}{2\sigma^2}\right)$$

besitzen. Die Zufallsvariable

$$V^2 = V_1^2 + V_2^2 + V_3^2$$

hat eine χ^2-Verteilung mit drei Freiheitsgraden. Ihre Dichte $g(v^2)$ ist nach Formel (9.4.4) durch

$$g(v^2) = \begin{cases} \dfrac{1}{\sigma^3 \sqrt{2\pi}}\, v \exp\left(-\dfrac{v^2}{2\sigma^2}\right) & \text{für} \quad v > 0, \\ 0 & \text{für} \quad v \leqq 0 \end{cases}$$

gegeben. Hieraus erhalten wir für die Geschwindigkeitsdichte $f(v)$ die *Maxwellsche Formel*

$$f(v) = \begin{cases} \dfrac{\sqrt{2}}{\sigma^3 \sqrt{\pi}} v^2 \exp\left(-\dfrac{v^2}{2\sigma}\right) & \text{für} \quad v > 0, \\ 0 & \text{für} \quad v \leqq 0. \end{cases} \tag{9.4.8}$$

Wir bestimmen die Dichte der Stichprobenfunktion

$$Y = \frac{1}{n} \sum_{k=1}^{n} X_k^2,$$

die gleich dem mittleren Quadrat von n unabhängigen Zufallsvariablen X_k ist, wobei alle X_k gleich normal $N(0; \sigma)$ verteilt sind.

Da $Y = \frac{1}{n} \chi^2$ gilt, ist die Dichte $f(y)$ der Zufallsvariablen Y durch

$$f(y) = \begin{cases} \dfrac{n^{\frac{n}{2}}}{2^{\frac{n}{2}} \sigma^n \Gamma\left(\dfrac{n}{2}\right)} e^{-\frac{ny}{2\sigma^2}} y^{\frac{n}{2}-1} & \text{für} \quad y > 0, \\ 0 & \text{für} \quad y \leqq 0 \end{cases} \tag{9.4.9}$$

bestimmt. Wir finden

$$E(Y) = \sigma^2, \qquad D^2(Y) = \frac{2\sigma^4}{n}. \tag{9.4.10}$$

Die Tafeln für χ^2 berücksichtigen gewöhnlich nur 30 Freiheitsgrade. FISHER [6] hat bewiesen, daß die Zufallsvariable $\sqrt{2\chi^2}$ asymptotisch normal $N\left(\sqrt{2n-1}; 1\right)$ verteilt ist, wenn die Anzahl n der Freiheitsgrade unbegrenzt wächst.

Für $n > 30$ kann man schon diese Grenzverteilung benutzen.

B. Nun seien die Zufallsvariablen X_k ($k = 1, 2, \ldots, n$) unabhängig, und es möge jedes X_k die Verteilung $N(m_k; 1)$ haben. Dann heißt die Stichprobenfunktion

$$\chi^2 = \sum_{k=1}^{n} X_k^2$$

das *nichtzentrale* χ^2. Die Dichte $h_n(u)$ dieser Stichprobenfunktion hat die Form

$$h_n(u) = \begin{cases} \dfrac{1}{\sqrt{\pi}\, 2^{n/2}} \exp\left[-\dfrac{1}{2}(\tau^2 + u)\right] u^{\frac{1}{2}n-1} \sum\limits_{j=0}^{\infty} \dfrac{(\tau^2 u)^j}{(2j)!} \dfrac{\Gamma\left(j + \frac{1}{2}\right)}{\Gamma\left(j + \frac{1}{2} n\right)} & \\ & \text{für} \quad u > 0, \\ 0 & \text{für} \quad u \leqq 0, \end{cases} \tag{9.4.11}$$

wobei $\tau^2 = \sum_{k=1}^{n} m_k^2$ ist. Man nennt τ^2 den *Nichtzentralitätsparameter* und n die *Anzahl der Freiheitsgrade.*

Den Beweis für die Formel (9.4.11) findet der Leser in der Monographie von ANDERSON [1]. Tafeln der nichtzentralen χ^2-Verteilung wurden von FIX [1] herausgegeben.

9.5. Die gemeinsame Verteilung der Stichprobenfunktionen $\bar{X}$ und S

Beispiel 9.5.1. Einer Grundgesamtheit, in der das Merkmal X normal $N(0; \sigma)$ verteilt ist, entnehmen wir eine einfache Stichprobe, die den Umfang n hat. Die beobachteten Werte der Zufallsvariablen X seien $x_1, x_2, \ldots, x_n$. Für diese Stichprobenwerte berechnen wir den Mittelwert $\bar{x}$ und die Standardabweichung s nach den Formeln

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k, \quad s^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2 = \frac{1}{n} \sum_{k=1}^{n} x_k^2 - \bar{x}^2.$$

Entnehmen wir der betrachteten Grundgesamtheit eine Serie von einfachen Stichproben, die alle den Umfang n haben, so werden wir im allgemeinen eine Serie von verschiedenen Wertepaaren $\bar{x}$ und s erhalten. Die von uns beobachteten Werte x und s stellen einfach den beobachteten Wert der zweidimensionalen Zufallsvariablen (X, S) dar. Ebenso sind die beobachteten Werte $\bar{x}$ und s, für sich betrachtet, Realisierungen der Zufallsvariablen $\bar{X}$ bzw. S, wobei

$$\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k, \quad S^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \bar{X})^2 \tag{9.5.1}$$

gesetzt ist.

In den Anwendungen ergibt sich häufig die folgende Aufgabe: Man hat in der Stichprobe die Werte $\bar{x}$ und s beobachtet, wobei es sich zeigte, daß $a \leqq \bar{x} < b$ und $s \geqq c$ ist. Wir suchen die Wahrscheinlichkeit für diese Ungleichungen, d. h., wir wollen feststellen, mit welcher angenäherten Häufigkeit die Werte $\bar{x}$ der Stichprobenfunktion $\bar{X}$ und die Werte s der Stichprobenfunktion S auftreten, die diese Ungleichungen erfüllen, wenn wir eine lange Serie von Stichproben machen.

Um diese Frage beantworten zu können, müssen wir die gemeinsame Verteilung der Stichprobenfunktion $(\bar{X}, S)$ finden.

A. Es seien X_k $(k = 1, 2, \ldots, n)$ unabhängige Zufallsvariable mit der gleichen Verteilung. Die Dichte ihrer Verteilung sei

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Wie wir schon in 9.4 bemerkten, bedeutet die Annahme $E(X_k) = 0$ keine Einschränkung der Allgemeinheit. Es sei $f(x_1, x_2, \ldots, x_n)$ die Dichte der n-dimensionalen Zufallsvariablen $(X_1, X_2, \ldots, X_n)$. Unter einem *Wahrscheinlichkeitselement* dieser Zufallsvariablen verstehen wir den Ausdruck

$$dP = f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n.$$

Da die Zufallsvariablen X_k unabhängig sind, erhalten wir

$$dP = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{n} x_k^2\right) dx_1\, dx_2 \cdots dx_n$$

$$= \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left(-\frac{n\overline{x}^2 + n s^2}{2\sigma^2}\right) dx_1 dx_2 \cdots dx_n. \tag{9.5.2}$$

Wir führen die Substitution

$$x_k = \overline{x} + s z_k \qquad (k = 1, 2, \ldots, n) \tag{9.5.3}$$

durch. Aus den Beziehungen

$$\sum_{k=1}^{n} x_k = n\overline{x}, \qquad \sum_{k=1}^{n} x_k^2 = n\overline{x}^2 + n s^2$$

folgen zwei Gleichungen, die zwischen den Veränderlichen z_k bestehen, nämlich

$$\sum_{k=1}^{n} z_k = 0, \qquad \sum_{k=1}^{n} z_k^2 = n. \tag{9.5.4}$$

Folglich sind zwei Veränderliche, zum Beispiel z_{n-1} und z_n, Funktionen der übrigen. Aus den Formeln (9.5.4) ergibt sich, daß entweder

$$z_{n-1} = \frac{A - B}{2}, \qquad z_n = \frac{A + B}{2} \tag{9.5.5}$$

oder

$$z_{n-1} = \frac{A + B}{2}, \qquad z_n = \frac{A - B}{2} \tag{9.5.6}$$

ist, wobei

$$A = -\sum_{k=1}^{n-2} z_k, \qquad B = \sqrt{2n - 3\sum_{k=1}^{n-2} z_k^2 - \sum_{\substack{k,j=1\\k\neq j}}^{n-2} z_k z_j}$$

gilt. Die Transformation (9.5.3) ist also nicht eineindeutig; jedem System $(z_1, z_2, \ldots, z_{n-2}, \overline{x}, s)$, wobei

$$s > 0, \qquad \sum_{k=1}^{n-2} z_k \neq 0 \quad \text{und} \quad \sum_{k=1}^{n-2} z_k^2 < n,$$

entsprechen zwei Systeme $(x_1, x_2, \ldots, x_n)$, nämlich das System

$$\begin{aligned} x_k &= \bar{x} + s z_k \qquad (k = 1, 2, \ldots, n-2), \\ x_{n-1} &= \bar{x} + s\,\frac{A-B}{2}, \\ x_n &= \bar{x} + s\,\frac{A+B}{2} \end{aligned} \tag{9.5.7}$$

und das System

$$\begin{aligned} x_k &= \bar{x} + s z_k \qquad (k = 1, 2, \ldots, n-2), \\ x_{n-1} &= \bar{x} + s\,\frac{A+B}{2}, \\ x_n &= \bar{x} + s\,\frac{A-B}{2}. \end{aligned} \tag{9.5.8}$$

Wir bemerken, daß die Funktionaldeterminanten beider Transformationen den gleichen absoluten Wert haben, da die eine Transformation aus der anderen durch Vertauschung von x_{n-1} und x_n hervorgeht. Wir erhalten für die Transformation (9.5.7)

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial \bar{x}} & \cdots & \frac{\partial x_n}{\partial \bar{x}} \\ \frac{\partial x_1}{\partial s} & \cdots & \frac{\partial x_n}{\partial s} \\ \frac{\partial x_1}{\partial z_1} & \cdots & \frac{\partial x_n}{\partial z_1} \\ \cdots & \cdots & \cdots \\ \frac{\partial x_1}{\partial z_{n-2}} & \cdots & \frac{\partial x_n}{\partial z_{n-2}} \end{vmatrix} = \begin{vmatrix} 1 & 1 & \ldots & 1 & 1 & 1 \\ z_1 & z_2 & \ldots & z_{n-2} & \frac{A-B}{2} & \frac{A+B}{2} \\ s & 0 & \ldots & 0 & \frac{s}{2}\left(\frac{\partial A}{\partial z_1} - \frac{\partial B}{\partial z_1}\right) & \frac{s}{2}\left(\frac{\partial A}{\partial z_1} + \frac{\partial B}{\partial z_1}\right) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & s & \frac{s}{2}\left(\frac{\partial A}{\partial z_{n-2}} - \frac{\partial B}{\partial z_{n-2}}\right) & \frac{s}{2}\left(\frac{\partial A}{\partial z_{n-2}} + \frac{\partial B}{\partial z_{n-2}}\right) \end{vmatrix}.$$

Nach einigen elementaren Umformungen finden wir

$$|J| = k s^{n-2}.$$

Dabei ist $k = k(z_1, \ldots, z_{n-2})$ eine gewisse Funktion, die nicht von $\bar{x}$ und s abhängt. Wenn wir noch beachten, daß die Dichte auf der rechten Seite von (9.5.2) den gleichen Wert für beide Transformationen (9.5.7) und (9.5.8) hat, und auf jede

dieser Transformationen die Formel (2.9.3) anwenden, dann erhalten wir aus (9.5.2)

$$\begin{aligned} dP &= g(\overline{x}, s, z_1, \ldots, z_{n-2})\, d\overline{x}\, ds\, dz_1 \cdots dz_{n-2} \\ &= 2\, \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left(-\frac{n\overline{x}^2 + n s^2}{2\sigma^2}\right) s^{n-2} k\, d\overline{x}\, ds\, dz_1 \cdots dz_{n-2}. \end{aligned} \tag{9.5.9}$$

Dabei ist

$$g(\overline{x}, s, z_1, \ldots, z_{n-2})$$

die Dichte der Zufallsvariablen

$$(\overline{X}, S, Z_1, \ldots, Z_{n-2}).$$

Ein Vergleich mit dem Beweis der Formel (2.4.14) erleichtert dem Leser das Verständnis für die Herleitung der Formel (9.5.9).

Wir wollen nun (9.5.9) in der folgenden Gestalt schreiben:

$$\begin{aligned} &g(\overline{x}, s, z_1, \ldots, z_{n-2})\, d\overline{x}\, ds\, dz_1 \cdots dz_{n-2} \\ &= \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n\overline{x}^2}{2\sigma^2}\right) d\overline{x} \cdot C\, ds \cdot \frac{\Gamma\left(\frac{n-1}{2}\right)}{n^{\frac{n}{2}}\, \pi^{\frac{n-1}{2}}} \cdot k(z_1, \ldots, z_{n-2})\, dz_1 \cdots dz_{n-2} \end{aligned}$$

mit

$$C = \frac{n^{\frac{n-1}{2}} s^{n-2} \exp\left(-\frac{n s^2}{2\sigma^2}\right)}{2^{\frac{n-3}{2}}\, \Gamma\left(\frac{n-1}{2}\right) \sigma^{n-1}}.$$

Hiermit ist das Wahrscheinlichkeitselement der Zufallsvariablen

$$(\overline{X}, S, Z_1, \ldots, Z_{n-2})$$

gleich dem Produkt dreier Faktoren. Der erste Faktor ist das Wahrscheinlichkeitselement der Zufallsvariablen $\overline{X}$, der zweite das der Zufallsvariablen S, während der dritte das Wahrscheinlichkeitselement der Zufallsvariablen $(Z_1, \ldots, Z_{n-2})$ ist. Somit sind nach (2.8.5) die Zufallsvariablen $\overline{X}$, S und $(Z_1, \ldots, Z_{n-2})$ unabhängig. Wenn wir mit $h(\overline{x}, s)$ die Dichte der Zufallsvariablen $(\overline{X}, S)$ bezeichnen, erhalten wir

$$h(\overline{x}, s) = \begin{cases} \dfrac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{n\overline{x}^2}{2\sigma^2}\right) \cdot \dfrac{n^{\frac{n-1}{2}} s^{n-2} \exp\left(-\dfrac{n s^2}{2\sigma^2}\right)}{2^{\frac{n-3}{2}}\, \Gamma\left(\dfrac{n-1}{2}\right) \sigma^{n-1}} & (s > 0), \\ 0 & (s \leqq 0). \end{cases} \tag{9.5.10}$$

B. Wir wollen jetzt die Verteilung der Stichprobenfunktion $U = S^2$ finden. Zu diesem Zweck substituieren wir im zweiten Faktor auf der rechten Seite von (9.5.10) $u = s^2$. Da $s > 0$ ist, ist die Transformation $u = s^2$ eineindeutig, und nach (2.4.19) ist die Dichte der Stichprobenfunktion U gleich

$$f_1(u) = \begin{cases} \dfrac{n^{\frac{n-1}{2}} u^{\frac{n-3}{2}} e^{-\frac{nu}{2\sigma^2}}}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right) \sigma^{n-1}} & \text{für} \quad u > 0, \\ 0 & \text{für} \quad u \leqq 0. \end{cases} \tag{9.5.11}$$

Wir betrachten jetzt die Zufallsvariable $Z = nS^2 = nU$. Wir bezeichnen ihre Dichtefunktion mit $f_2(z)$ und erhalten

$$f_2(z) = \begin{cases} \dfrac{z^{\frac{n-3}{2}} e^{-\frac{z}{2\sigma^2}}}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right) \sigma^{n-1}} & \text{für} \quad z > 0, \\ 0 & \text{für} \quad z \leqq 0. \end{cases} \tag{9.5.12}$$

Vergleichen wir (9.5.12) mit der Formel (9.4.4), die uns die Dichte der Zufallsvariablen

$$\chi^2 = \sum_{k=1}^{n} X_k^2$$

angibt, so sehen wir, daß die Zufallsvariable Z dieselbe Verteilung wie die Zufallsvariable χ^2 mit $n - 1$ Freiheitsgraden hat. Die Formel (9.5.12) kann man nämlich aus (9.4.4) erhalten, indem man überall n durch $n - 1$ ersetzt. Dieses Ergebnis stimmt mit der Anschauung überein; denn die Zufallsvariable χ^2 ist nach Definition eine Summe von n unabhängigen Zufallsvariablen, während

$$nS^2 = \sum_{k=1}^{n} (X_k - \overline{X})^2$$

eine Summe von n Zufallsvariablen ist, die der Gleichung

$$\sum_{k=1}^{n} X_k = n\overline{X}$$

genügen. Durch dieses Ergebnis können wir den Begriff des Freiheitsgrades besser verstehen.

Aus (9.4.5) erhalten wir

$$E(nS^2) = (n-1)\,\sigma^2, \qquad D^2(nS^2) = 2(n-1)\,\sigma^4.$$

Hieraus folgen die Gleichungen

$$E(S^2) = \frac{n-1}{n}\sigma^2, \quad D^2(S^2) = \frac{2(n-1)}{n^2}\sigma^4. \tag{9.5.13}$$

C. Wir wenden uns nochmals dem Ausdruck (9.5.10) zu und sehen: Haben die unabhängigen Zufallsvariablen X_k die gleiche Normalverteilung, dann ist die gemeinsame Dichte der Zufallsvariablen $\overline{X}$ und S gleich dem Produkt der Dichten dieser Zufallsvariablen. Diese Zufallsvariablen sind also nach (2.8.5) unabhängig. Dieses äußerst wichtige Ergebnis stammt von FISHER [5]. Es hat folgende praktische Bedeutung: Liegt eine Serie von unabhängigen einfachen Stichproben vor, die alle aus derselben normalen Gesamtheit stammen, und teilen wir diese Serie so in Gruppen ein, daß ein und derselben Gruppe alle diejenigen Stichproben angehören, für die die berechneten s-Werte gleich (oder angenähert gleich) sind, dann ist in jeder dieser Gruppen die Häufigkeitsverteilung der $\bar{x}$-Werte, die zu irgendeinem festgelegten Intervall gehören, in Annäherung dieselbe. Ebenso kann man die Stichproben hinsichtlich der $\bar{x}$-Werte gruppieren; auch hier ist die Häufigkeitsverteilung der s-Werte, die zu irgendeinem festgelegten Intervall gehören, in jeder dieser Gruppen angenähert die gleiche.

Auch der umgekehrte Satz erwies sich als richtig: *Sind die Stichprobenfunktionen $\overline{X}$ und S unabhängig, dann sind die Zufallsvariablen X_k normal verteilt.* Den Nachweis lieferten GEARY [1], LUKACS [1], KAWATA und SAKAMOTO [1] und ZINGER [1]. Später haben LUKACS [2] sowie BASU und LAHA [1] diesen Satz verallgemeinert.

Beispiel 9.5.2. Einer Gesamtheit, deren Merkmal X normal $N(1; 2)$ verteilt ist, entnehmen wir eine einfache Stichprobe, die 12 Elemente enthält. Dabei werden die folgenden Werte der Zufallsvariablen X beobachtet:

$$x_1 = 2{,}0, \quad x_2 = 2{,}5, \quad x_3 = 0{,}5, \quad x_4 = 1{,}0, \quad x_5 = 0{,}0, \quad x_6 = -0{,}9,$$

$$x_7 = 5{,}1, \quad x_8 = -1{,}5, \quad x_9 = 0{,}8, \quad x_{10} = 1{,}1, \quad x_{11} = 0{,}8, \quad x_{12} = 0{,}4.$$

Hieraus erhalten wir

$$\bar{x} = \frac{1}{12}\sum_{k=1}^{12} x_k = 0{,}98, \qquad s^2 = \frac{1}{12}\sum_{k=1}^{12} x_k^2 - \bar{x}^2 = 2{,}69.$$

Es gilt also $z = ns^2 = 12s^2 = 32{,}28$. Wie groß ist die Wahrscheinlichkeit dafür, daß die Zufallsvariable Z nicht kleiner als der beobachtete Wert $z = 32{,}28$ ist?

Wie wir wissen, hat die Zufallsvariable Z eine χ^2-Verteilung mit 11 Freiheitsgraden. Berücksichtigen wir, daß die Standardabweichung der Zufallsvariablen X gleich 2 ist, so können wir die Wahrscheinlichkeit $P(Z \geqq 32{,}28)$ für 11 Freiheitsgrade aus den Tafeln für die χ^2-Verteilung ablesen. Dabei interpolieren wir wie in Beispiel 9.4.2:

$$P(Z \geqq 32{,}28) = P\left(\frac{nS^2}{\sigma^2} \geqq \frac{32{,}28}{4}\right) = P(\chi^2 \geqq 8{,}07) \approx 0{,}70.$$

9.6. Die Studentsche t-Verteilung

A. In 9.4 betrachteten wir die Verteilung der durch

$$\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$$

gegebenen Stichprobenfunktion unter der Voraussetzung, daß die Zufallsvariablen X_k $(k = 1, 2, \ldots, n)$ unabhängig sind und die gleiche Normalverteilung $N(m; \sigma)$ besitzen. Dabei stellten wir fest, daß $\overline{X}$ normal $N\left(m; \frac{\sigma}{\sqrt{n}}\right)$ verteilt ist. Die Verteilung der Zufallsvariablen $\overline{X}$ ist also für ein gegebenes m unbekannt, wenn σ unbekannt ist. Hier können wir offenbar nicht den aus der Stichprobe errechneten Wert s als σ nehmen, denn S ist eine Zufallsvariable, die im allgemeinen in verschiedenen Stichproben verschiedene Werte annimmt. Um den Parameter m bei unbekanntem σ abschätzen zu können (vgl. Kap. 13), ist es notwendig, eine Stichprobenfunktion zu betrachten, die eine Funktion von m ist und deren Verteilung nicht von σ abhängt. Dieses Problem löste GOSSET (er schrieb unter dem Pseudonym STUDENT), indem er eine Stichprobenfunktion einführte, die wir mit t bezeichnen und nach STUDENT benennen. Wir definieren sie folgendermaßen:

Es seien X_k $(k = 1, 2, \ldots, n)$ unabhängige Zufallsvariable mit der gleichen Normalverteilung $N(m; \sigma)$. Wir definieren[1])

$$t = \frac{\overline{X} - m}{S} \sqrt{n-1}, \tag{9.6.1}$$

wobei die Zufallsvariablen $\overline{X}$ und S durch die Formeln (9.5.1) bestimmt sind.

Wir bestimmen zuerst die Dichte $f(v)$ der Zufallsvariablen $V = \dfrac{\overline{X} - m}{S}$.

Aus (9.5.10) folgt, daß diese Zufallsvariable V der Quotient zweier unabhängiger Zufallsvariabler ist und daß die Dichte des Zählers bzw. die des Nenners die Form

$$f_1(\overline{x} - m) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n(\overline{x} - m)^2}{2\sigma^2}\right)$$

bzw.

$$f_2(s) = \frac{n^{\frac{n-1}{2}} s^{n-2} e^{-\frac{ns^2}{2\sigma^2}}}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-1}{2}\right) \sigma^{n-1}} \tag{9.6.2}$$

[1]) Aus traditionellen Gründen wollen wir im Gegensatz zu der sonst von uns benutzten Bezeichnungsweise die Studentsche Zufallsvariable t und ihre Werte mit demselben kleinen Buchstaben t bezeichnen.

haben. Wir erhalten also aus (2.9.16') die Dichte der Zufallsvariablen V in der Form

$$f(v) = \int_0^\infty \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{nv^2s^2}{2\sigma^2}} \frac{n^{\frac{n-1}{2}} s^{n-2} e^{-\frac{ns^2}{2\sigma^2}}}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-1}{2}\right) \sigma^{n-1}} s\,ds$$

$$= \frac{n^{\frac{n}{2}}}{\sigma^n \sqrt{\pi}\, 2^{\frac{n-2}{2}} \Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty e^{-\frac{ns^2(v^2+1)}{2\sigma^2}} s^{n-1}\,ds.$$

Im Integral substituieren wir $z = s^2$ und erhalten

$$f(v) = \frac{n^{\frac{n}{2}}}{\sigma^n \sqrt{\pi}\, 2^{\frac{n}{2}} \Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty e^{-\frac{nz(v^2+1)}{2\sigma^2}} z^{\frac{n-2}{2}}\,dz.$$

Wenn wir in (5.8.5)

$$a = \frac{n(v^2+1)}{2\sigma^2} \quad \text{und} \quad p = \frac{n}{2}$$

setzen und beachten, daß $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ ist, erhalten wir nach einigen Umformungen

$$f(v) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{(v^2+1)^{\frac{n}{2}}} \qquad (-\infty < v < \infty). \tag{9.6.3}$$

Man kann den Ausdruck (9.6.3) in der Form

$$f(v) = \frac{1}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)} \frac{1}{(v^2+1)^{\frac{n}{2}}} \qquad (-\infty < v < \infty) \tag{9.6.4}$$

schreiben. Aus (9.6.4) ergibt sich, daß die Dichte $g(t)$ der durch (9.6.1) gegebenen Zufallsvariablen t die Form

$$g(t) = \frac{1}{\sqrt{n-1}\, B\left(\frac{1}{2}, \frac{n-1}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{\frac{n}{2}}} \qquad (-\infty < t < \infty) \tag{9.6.5}$$

hat.

Die Dichte der Studentschen Zufallsvariablen t ist also nicht von σ abhängig. Diese Tatsache hat, wie wir schon oben erwähnten, viele Anwendungen der t-Verteilung zur Folge. Wir wollen sagen, daß (9.6.5) die *Studentsche t-Verteilung mit* $n-1$ *Freiheitsgraden* bestimmt. Die Dichte der Zufallsvariablen t ist symmetrisch in bezug auf den Punkt $t = 0$.

Für die Studentsche t-Verteilung existieren nur die Momente der Ordnung $k < n-1$. So existieren zum Beispiel für $n = 1$ überhaupt keine Momente. Wie der Leser nachprüfen möge, ist für $n = 2$ die Studentsche t-Verteilung ein

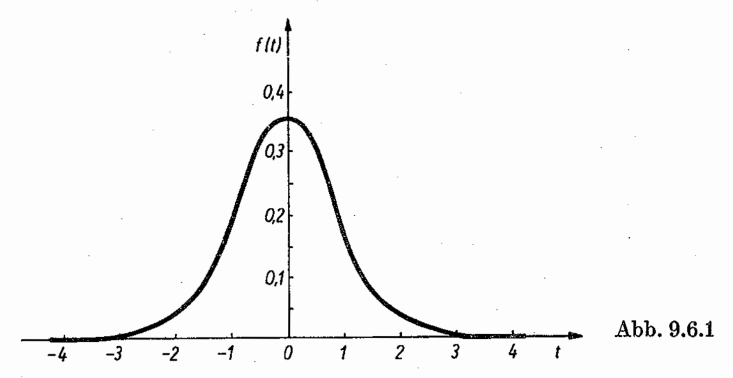


Abb. 9.6.1

Sonderfall der Cauchyschen Verteilung, die, wie wir wissen (vgl. 5.10), keine endlichen Momente besitzt.

Die Wahrscheinlichkeit, daß t im Intervall (t_1, t_2) enthalten ist, wird durch das Integral

$$P(t_1 < t < t_2) = \frac{1}{\sqrt{n-1}\, B\left(\frac{1}{2}, \frac{n-1}{2}\right)} \int_{t_1}^{t_2} \frac{dt}{\left(1 + \frac{t^2}{n-1}\right)^{\frac{n}{2}}}$$

ausgedrückt. Abb. 9.6.1 stellt die Studentsche t-Dichte für $n = 3$ dar.

Denken wir an den Beweis der Formel (9.6.5) und definieren wir die Zufallsvariable U durch

$$U = \frac{Z}{\sqrt{W/r}}, \tag{9.6.1'}$$

wobei Z und W unabhängige Zufallsvariable sind, Z die Normalverteilung $N(0; 1)$ und W eine χ^2-Verteilung mit r Freiheitsgraden besitzt, so sehen wir leicht, daß U die Studentsche t-Verteilung mit r Freiheitsgraden hat.

B. Vergleichen wir die Tafeln der Studentschen t-Verteilung mit 30 Freiheitsgraden mit der entsprechenden Tafel der Normalverteilung, so stellen wir fest, daß beide Teile beinahe identisch sind. Dies ist deshalb so, weil die Studentsche

t-Verteilung ziemlich schnell gegen die Normalverteilung konvergiert, wenn die Anzahl der Freiheitsgrade ins Unendliche wächst. Es gilt nämlich der folgende Satz:

Satz 9.6.1. *Die Folge $\{F_n(t)\}$ der Studentschen t-Verteilungsfunktionen mit n Freiheitsgraden genügt der Beziehung*

$$\lim_{n\to\infty} F_n(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{t^2}{2}}\, dt. \tag{9.6.6}$$

Beweis. Wir schreiben die Formel (9.6.1) in der Gestalt

$$t_n = \frac{\dfrac{\overline{X} - m}{\sigma}\sqrt{n}}{\sqrt{\dfrac{nS^2}{(n-1)\sigma^2}}} = \frac{Y_n}{\sqrt{Z_n}} = \frac{Y_n}{V_n}.$$

Wie wir wissen, ist für jedes natürliche n die Zufallsvariable Y_n normal $N(0; 1)$ verteilt. Wir beweisen, daß die Folge $\{V_n\}$ stochastisch gegen 1 konvergiert.

Tatsächlich hat die Zufallsvariable nS^2 eine Gammaverteilung mit $p = \dfrac{n-1}{2}$ und $b = \dfrac{1}{2\sigma^2}$. Die charakteristische Funktion $\varphi_n(\alpha)$ der Zufallsvariablen Z_n ist somit nach (5.8.8) gleich

$$\varphi_n(\alpha) = \left(1 - \frac{i\alpha}{\dfrac{n-1}{2}}\right)^{-\frac{n-1}{2}}.$$

Daraus erhalten wir

$$\lim_{n\to\infty} \varphi_n(\alpha) = e^{i\alpha}.$$

Aus der letzten Gleichung folgt, daß die Folge $\{Z_n\}$ und somit auch die Folge $\{V_n\}$ stochastisch gegen 1 konvergiert. Aus der Behauptung δ) des Satzes 6.14.1 erhalten wir die Gleichung (9.6.6).

C. Allgemeiner als die Studentsche t-Stichprobenfunktion ist die *nichtzentrale t-Stichprobenfunktion*. Diese Stichprobenfunktion definieren wir folgendermaßen:

Es seien X_k $(k = 1, 2, \ldots, n)$ unabhängige Zufallsvariable mit der gleichen Normalverteilung $N(m; \sigma)$. Als *nichtzentrales t* bezeichnen wir den Ausdruck

$$Y = \frac{\overline{X} - x_0}{S}\sqrt{n-1};$$

dabei sind $\bar{X}$ und S durch die Formeln (9.5.1) bestimmt, während x_0 irgendein konstanter Wert ist.

Die nichtzentrale t-Stichprobenfunktion unterscheidet sich vom Studentschen t dadurch, daß an Stelle des Mittelwertes der Zufallsvariablen X im Zähler ein beliebiger Wert x_0 auftritt. Es ist klar, daß für $x_0 = m$ die Zufallsvariable Y mit der Studentschen t-Stichprobenfunktion übereinstimmt.

Die Herleitung der Dichte der Zufallsvariable Y ist ähnlich der der Formel (9.6.5). Diese Dichte hat die Form

$$f(y) = \frac{n^{\frac{n}{2}}}{\sqrt{n-1}\,\Gamma\left(\frac{n-1}{2}\right)\sqrt{\pi}\,2^{\frac{n}{2}-1}}\, e^{-\frac{n(n-1)Q^2}{2(y^2+n-1)}} \times \frac{1}{\left(1+\frac{y^2}{n-1}\right)^{\frac{n}{2}}} \int\limits_{\frac{-Qy}{\sqrt{y^2+n-1}}}^{\infty} e^{-\frac{n}{2}u^2}\left[u+\frac{Qy}{\sqrt{y^2+n-1}}\right]^{n-1} du. \tag{9.6.7}$$

Dabei ist Q ein durch

$$Q = \frac{m - x_0}{\sigma}$$

bestimmter Parameter. Ist $x_0 = m$, also $Q = 0$, so geht (9.6.7) in (9.6.5) über.

Umfangreiche Tabellen für die nichtzentrale t-Verteilung haben Resnikoff und Liebermann [1] herausgegeben; Harley [1] sowie Merrington und Pearson [1] haben eine Näherungsformel für diese Verteilung angegeben.

D. Wir geben noch eine Stichprobenfunktion an, die dieselbe Verteilung wie die Studentsche t-Stichprobenfunktion hat und große Bedeutung für die Anwendungen besitzt.

Es seien $X_1, X_2, \ldots, X_{n_1}$ und $Y_1, Y_2, \ldots, Y_{n_2}$ unabhängige Zufallsvariable mit derselben Normalverteilung $N(m; \sigma)$. Wir setzen

$$\bar{X} = \frac{1}{n_1}\sum_{k=1}^{n_1} X_k, \quad \bar{Y} = \frac{1}{n_2}\sum_{l=1}^{n_2} Y_l,$$

$$S_1^2 = \frac{1}{n_1}\sum_{k=1}^{n_1}(X_k - \bar{X})^2, \quad S_2^2 = \frac{1}{n_2}\sum_{l=1}^{n_2}(Y_l - \bar{Y})^2.$$

Wie bekannt, haben $\bar{X}$ bzw. $\bar{Y}$ die Normalverteilungen

$$N\left(m; \frac{\sigma}{\sqrt{n_1}}\right) \quad \text{bzw.} \quad N\left(m; \frac{\sigma}{\sqrt{n_2}}\right).$$

Die Zufallsvariable $\dfrac{\bar{X}-\bar{Y}}{\sigma}$ hat folglich die Verteilung

$$N\left(0;\ \sigma\sqrt{\frac{n_1+n_2}{n_1 n_2}}\right).$$

Daraus folgt, daß die Zufallsvariable

$$Z = \frac{\bar{X}-\bar{Y}}{\sigma}\sqrt{\frac{n_1 n_2}{n_1+n_2}} \tag{9.6.8}$$

die Verteilung $N(0; 1)$ hat. Die Zufallsvariable

$$W = \frac{n_1 S_1^2 + n_2 S_2^2}{\sigma^2} \tag{9.6.9}$$

hat eine χ^2-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden. Dies folgt aus dem Additionssatz für χ^2 und daraus, daß die Zufallsvariablen S_1^2 und S_2^2 unabhängig sind.

Wir betrachten die Zufallsvariable

$$\begin{aligned} U &= \frac{\bar{X}-\bar{Y}}{\sqrt{n_1 S_1^2 + n_2 S_2^2}}\sqrt{\frac{n_1 n_2}{n_1+n_2}(n_1+n_2-2)} \\ &= \frac{Z}{\sqrt{W/(n_1+n_2-2)}}, \end{aligned} \tag{9.6.10}$$

wobei Z bzw. W durch die Formeln (9.6.8) bzw. (9.6.9) definiert sind. Vergleichen wir (9.6.10) und (9.6.1'), so sehen wir, daß U die Studentsche t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden hat. Weiterhin sehen wir, daß die Verteilung der Zufallsvariablen U sowohl von m als auch von σ unabhängig ist. Dieses Ergebnis stammt von FISHER [5].

E. Eine Verallgemeinerung der Studentschen t-Verteilung (oder genauer der t^2-Verteilung) auf mehrdimensionale Zufallsvariable ist die T^2-Verteilung von HOTELLING [2].

Es seien die $X_k = (X_{k1}, X_{k2}, \ldots, X_{kr})$ $(k = 1, 2, \ldots, n)$ unabhängige r-dimensionale Zufallsvektoren mit gleicher Normalverteilung. Ferner sei $m_j = E(X_{kj})$ und $\lambda_{ij} = E[(X_{kj} - m_j)(X_{ki} - m_i)]$ $(i, j = 1, 2, \ldots, r)$, wobei die Matrix M der Varianzen und Kovarianzen die Determinante $|M| \neq 0$ hat.

Wir setzen ferner

$$\bar{X}_j = \frac{1}{n}\sum_{k=1}^{n} X_{kj}, \tag{9.6.11}$$

$$W_{ji} = \frac{1}{n} \sum_{k=1}^{n} (X_{kj} - \overline{X}_j)(X_{ki} - \overline{X}_i), \tag{9.6.12}$$

$$Q = \begin{pmatrix} W_{11} \dots W_{1r} \\ \dots\dots\dots \\ W_{r1} \dots W_{rr} \end{pmatrix}. \tag{9.6.13}$$

Die W_{ji} nennen wir die *Stichprobenvarianzen* (für $j = i$) sowie die *Stichprobenkovarianzen* (für $j \neq i$), und Q bezeichnen wir als die *Matrix der Stichprobenmomente zweiter Ordnung.*

Den Ausdruck

$$T^2 = (n-1) \sum_{j,i=1}^{r} \frac{|Q_{ji}|}{|Q|} (\overline{X}_j - m_j)(\overline{X}_i - m_i), \tag{9.6.14}$$

wobei $|Q|$ die mit der Wahrscheinlichkeit 1 von 0 verschiedene Determinante der Matrix Q und $|Q_{ji}|$ das algebraische Komplement zu W_{ji} in der Determinante der Matrix Q ist, nennen wir die *Hotellingsche T^2-Verteilung.*

Die Dichtefunktion $g(y)$ der Variablen T^2 ist definiert durch

$$g(y) = \begin{cases} \dfrac{1}{B\left(\dfrac{n-r}{2}, \dfrac{r}{2}\right)(n-1)} \dfrac{\left(\dfrac{y}{n-1}\right)^{r/2-1}}{\left(1 + \dfrac{y}{n-1}\right)^{n/2}} & \text{für} \quad y > 0, \\ 0 & \text{für} \quad y \leqq 0. \end{cases} \tag{9.6.15}$$

Den Beweis der Formel (9.6.15) findet der Leser bei SCHMETTERER [1] oder bei CRAMÉR [2].

Es sei bemerkt, daß die Dichtefunktion (9.6.15) von den Momenten zweiter Ordnung λ_{ij} der Zufallsvektoren $(X_{k1}, X_{k2}, \dots, X_{kr})$ unabhängig ist. Wir sagen, T^2 habe $n - 1$ *Freiheitsgrade.*

Die Hotellingsche T^2-Verteilung spielt in der mathematischen Statistik im Fall der Beobachtung von Zufallsvektoren die gleiche Rolle wie die Studentsche t-Verteilung im Fall der Beobachtung eindimensionaler Zufallsvariabler.

9.7. Die Fishersche Z-Verteilung

Es seien X_k $(k = 1, 2, \dots, n_1)$ und Y_l $(l = 1, 2, \dots, n_2)$ unabhängige Zufallsvariable mit derselben Normalverteilung $N(0; \sigma)$. Wir setzen

$$S_1^2 = \frac{1}{n_1} \sum_{k=1}^{n_1} (X_k - \overline{X})^2, \quad S_2^2 = \frac{1}{n_2} \sum_{l=1}^{n_2} (Y_l - \overline{Y})^2$$

und bilden die Stichprobenfunktion

$$U = \frac{S_1}{S_2}. \tag{9.7.1}$$

Nach Voraussetzung sind die Zufallsvariablen S_1 und S_2 unabhängig, und ihre Dichten sind nach (9.5.10) durch

$$f_i(s_i) = \begin{cases} \dfrac{n_i^{\frac{n_i-1}{2}} s_i^{n_i-2} e^{\frac{-n_i s_i^2}{2\sigma^2}}}{2^{\frac{n_i-3}{2}} \Gamma\left(\dfrac{n_i-1}{2}\right) \sigma^{n_i-1}} & \text{für} \quad s_i > 0, \\ 0 & \text{für} \quad s_i \leqq 0 \end{cases}$$

$(i = 1, 2)$ bestimmt. Nach (2.9.16′) ist die Dichte der Zufallsvariablen U durch die Funktion

$$f(u) = \begin{cases} C_{n_1,n_2} \displaystyle\int_0^\infty s_2^{n_1+n_2-3} u^{n_1-2} \exp\left[-\frac{s_2^2}{2\sigma^2}(n_1 u^2 + n_2)\right] ds_2 & \text{für} \quad u > 0 \\ \quad \text{mit} \quad C_{n_1,n_2} = \dfrac{n_1^{\frac{n_1-1}{2}} n_2^{\frac{n_2-1}{2}}}{2^{\frac{n_1+n_2}{2}-3} \Gamma\left(\dfrac{n_1-1}{2}\right) \Gamma\left(\dfrac{n_2-1}{2}\right) \sigma^{n_1+n_2-2}}, & \\ 0 & \text{für} \quad u \leqq 0 \end{cases} \tag{9.7.2}$$

dargestellt. Wir setzen $y = s_2^2$ und erhalten

$$f(u) = \begin{cases} \dfrac{1}{2} C_{n_1,n_2} u^{n_1-2} \displaystyle\int_0^\infty y^{\frac{n_1+n_2}{2}-2} \exp\left(-\frac{n_1 u^2 + n_2}{2\sigma^2} y\right) dy & \text{für} \quad u > 0, \\ 0 & \text{für} \quad u \leqq 0. \end{cases}$$

Wenn wir die Formel (5.8.5) benutzen, in der wir

$$a = \frac{n_1 u^2 + n_2}{2\sigma^2}, \quad p = \frac{n_1 + n_2}{2} - 1$$

setzen, so finden wir

$$f(u) = \begin{cases} \dfrac{n_1^{\frac{n_1-1}{2}} n_2^{\frac{n_2-1}{2}} \Gamma\left(\dfrac{n_1+n_2}{2} - 1\right)}{2^{\frac{n_1+n_2}{2}-2} \sigma^{n_1+n^2-2} \Gamma\left(\dfrac{n_1-1}{2}\right) \Gamma\left(\dfrac{n_2-1}{2}\right)} \dfrac{u^{n_1-2} (2\sigma^2)^{\frac{n_1+n_2}{2}-1}}{(n_1 u^2 + n_2)^{\frac{n_1+n_2}{2}-1}} & \text{für} \quad u > 0, \\ 0 & \text{für} \quad u \leqq 0. \end{cases}$$

Nach Einführung von $B(p, q)$ ergibt sich

$$f(u) = \begin{cases} \dfrac{2n_1^{\frac{n_1-1}{2}} n_2^{\frac{n_2-1}{2}}}{B\left(\dfrac{n_1-1}{2}, \dfrac{n_2-1}{2}\right)} \dfrac{u^{n_1-2}}{(n_1 u^2 + n_2)^{\frac{n_1+n_2}{2}-1}} & \text{für} \quad u > 0, \\ 0 & \text{für} \quad u \leqq 0. \end{cases} \tag{9.7.3}$$

Der Ausdruck (9.7.3) hängt nicht von σ ab. Wir setzen

$$u = \left[\frac{n_2(n_1 - 1)}{n_1(n_2 - 1)}\right]^{\frac{1}{2}} e^z \quad (-\infty < z < \infty). \tag{9.7.4}$$

Dann gilt

$$g(z) = \frac{2n_1^{\frac{n_1-1}{2}} n_2^{\frac{n_2-1}{2}}}{B\left(\dfrac{n_1-1}{2}, \dfrac{n_2-1}{2}\right)} \left[\frac{n_2(n_1-1)}{n_1(n_2-1)}\right]^{\frac{n_1-1}{2}} e^{z(n_1-1)} \frac{1}{\left(n_2 \dfrac{n_1-1}{n_2-1} e^{2z} + n_2\right)^{\frac{n_1+n_2}{2}-1}}$$

$$= \frac{2(n_1-1)^{\frac{n_1-1}{2}} (n_2-1)^{\frac{n_2-1}{2}}}{B\left(\dfrac{n_1-1}{2}, \dfrac{n_2-1}{2}\right)} \frac{e^{z(n_1-1)}}{[(n_1-1)e^{2z} + n_2 - 1]^{\frac{n_1+n_2}{2}-1}} \quad (-\infty < z < \infty).$$

Wenn wir $r_1 = n_1 - 1$, $r_2 = n_2 - 1$ setzen, erhalten wir

$$g(z) = \frac{2r_1^{\frac{r_1}{2}} r_2^{\frac{r_2}{2}}}{B\left(\dfrac{r_1}{2}, \dfrac{r_2}{2}\right)} \frac{e^{r_1 z}}{(r_1 e^{2z} + r_2)^{\frac{r_1+r_2}{2}}} \quad (-\infty < z < \infty). \tag{9.7.5}$$

Die Stichprobenfunktion Z, die mit der Stichprobenfunktion U durch die Formel (9.7.4) zusammenhängt, nennt man die *Fishersche Z-Stichprobenfunktion* (vgl. Fisher [7]). Das Zahlenpaar (r_1, r_2) bezeichnet man als *Freiheitsgrad* der Fisherschen Z-Stichprobenfunktion.

Die Zufallsvariable

$$F = \exp(2Z)$$

heißt *Snedecorsche F-Stichprobenfunktion*. Ihre Dichte läßt sich aus der Formel (9.7.5) herleiten.

Wir zeigen jetzt an einem Beispiel, wie die Fishersche Z-Verteilung in der Praxis angewandt wird.

Beispiel 9.7.1. Wir untersuchen eine Warenmenge. Das Merkmal X dieser Waren sei normal verteilt, die Standardabweichung dieser Verteilung aber unbekannt. Wir entnehmen zwei unabhängige einfache Stichproben. Die eine Probe habe den Umfang $n_1 = 5$, die zweite den Umfang $n_2 = 6$. Die Standardabweichung betrage in der ersten Stichprobe $s_1 = 1{,}3$ und in der zweiten $s_2 = 1$. Da beide Stichproben aus derselben Gesamtheit (Warenmenge) stammen, kann man erwarten, daß sich s_1 nur wenig von s_2 unterscheidet. Nun zeigte es sich jedoch, daß $\frac{s_1}{s_2} = 1{,}3$ ist. Wir fragen nach der Wahrscheinlichkeit dafür, daß in zwei Stichproben, die aus derselben normalen Gesamtheit stammen, das Verhältnis $\frac{S_1}{S_2}$ nicht kleiner als der beobachtete Wert 1,3 ist.

Diese Frage wollen wir mit Hilfe der Fisherschen Z-Stichprobenfunktion beantworten. Den Wert $u = 1{,}3$ betrachten wir als beobachteten Wert der Stichprobenfunktion U, die durch (9.7.1) bestimmt wurde. Aus (9.7.4) erhalten wir

$$z = \log \frac{u}{\sqrt{\frac{n_2(n_1-1)}{n_1(n_2-1)}}} = \log \frac{1{,}3}{\sqrt{0{,}96}} \approx 0{,}2826.$$

Dies ist der beobachtete Wert der Fisherschen Stichprobenfunktion Z, deren Freiheitsgrade durch das Zahlenpaar (4, 5) gegeben sind.

Wir stellen die Frage: Ist die Wahrscheinlichkeit von $U \geqq 1{,}3$ nicht kleiner als 0,05?

Aus der Tafel VI lesen wir den Wert z_0 ab, für den die Relation $P(Z \geqq z_0) = 0{,}05$ gilt; er beträgt $z_0 \approx 0{,}8236$. Da der von uns beobachtete z-Wert nur 0,2826 beträgt, können wir die gestellte Frage bejahend beantworten.

Bei Durchsicht des Beweises für die Formel (9.7.5) sehen wir, daß die durch

$$U = \frac{1}{2} \log \left(\frac{W_1}{r_1} : \frac{W_2}{r_2} \right)$$

definierte Zufallsvariable U eine Fishersche Z-Verteilung mit (r_1, r_2) Freiheitsgraden hat, wobei W_1 und W_2 unabhängige Zufallsvariable sind; W_i $(i = 1, 2)$ hat dabei eine χ^2-Verteilung mit r_i Freiheitsgraden.

Es möge W_1 eine nichtzentrale χ^2-Verteilung mit dem Nichtzentralitätsparameter τ^2 und r_1 Freiheitsgraden sein (vgl. 9.4.B), und W_2 habe eine χ^2-Verteilung mit r_2 Freiheitsgraden; ferner seien W_1 und W_2 unabhängig. Die Zufallsvariable

$$F = \frac{W_1}{r_1} : \frac{W_2}{r_2}$$

heißt die *nichtzentrale Snedecorsche Stichprobenfunktion* F, und entsprechend bezeichnet man den Ausdruck $Z = \frac{1}{2} \log F$ als die *nichtzentrale Fishersche*

Stichprobenfunktion Z. Die Dichtefunktion $g_1(z)$ der nichtzentralen Fisherschen Stichprobenfunktion Z läßt sich aus (2.9.16), (9.4.4) und (9.4.11) ermitteln und hat die Gestalt

$$g_1(z) = \frac{2 r_1^{r_1/2} r_2^{r_2/2}}{\Gamma\left(\frac{1}{2} r_2\right)} \exp\left(-\frac{1}{2}\tau^2 + r_1 z\right) \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2} e^{2z} \tau^2 r_1\right)^j \Gamma\left(\frac{1}{2}(r_1 + r_2) + j\right)}{j!\, \Gamma\left(\frac{1}{2} r_1 + j\right) (r_1 e^{2z} + r_2)^{(r_1+r_2)/2+j}}$$

Für $\tau^2 = 0$ geht die letzte Gleichung über in (9.7.5).

9.8. Die Verteilung von $\bar{X}$ in Stichproben aus einigen nichtnormalen Grundgesamtheiten

Bisher haben wir uns mit Stichprobenfunktionen befaßt, die Funktionen von unabhängigen Zufallsvariablen $X_1, \ldots, X_n$ waren, wobei alle X_k dieselbe Normalverteilung hatten. Jetzt wollen wir uns mit Stichprobenfunktionen beschäftigen, die über nichtnormalen Grundgesamtheiten definiert sind.

Die unabhängigen Zufallsvariablen X_k $(k = 1, 2, \ldots, j)$ mögen alle dieselbe *Binomialverteilung* haben, deren Wahrscheinlichkeitsfunktion für $r = 0, 1, \ldots, n$ durch

$$P(X_k = r) = \binom{n}{r} p^r q^{n-r}$$

gegeben ist $(0 < p < 1,\ q = 1 - p)$. Wir wollen die Verteilung des arithmetischen Mittels

$$\bar{X} = \frac{1}{j} \sum_{k=1}^{j} X_k \tag{9.8.1}$$

dieser Zufallsvariablen bestimmen.

Aus (5.2.3) folgt, daß die charakteristische Funktion der Zufallsvariablen X_k für jedes $k = 1, 2, \ldots, j$ durch

$$\varphi(t) = (q + p e^{it})^n$$

bestimmt ist. Nach (4.2.15) und (4.4.3) hat die charakteristische Funktion der Zufallsvariablen $\bar{X}$ die Form

$$\varphi_1(t) = \left(q + p e^{\frac{it}{j}}\right)^{nj}. \tag{9.8.2}$$

Dieser Ausdruck stellt die charakteristische Funktion einer Zufallsvariablen mit einer modifizierten Binomialverteilung dar; die Zufallsvariable $\bar{X}$ kann die Werte

$$0, \frac{1}{j}, \frac{2}{j}, \ldots, \frac{nj}{j} = n$$

mit den Wahrscheinlichkeiten

$$P\left(\bar{X} = \frac{l}{j}\right) = \binom{nj}{l} p^l q^{nj-l} \quad (l = 0, 1, \ldots, nj)$$

annehmen.

Wir betrachten nun die unabhängigen Zufallsvariablen

$$X_k \ (k = 1, 2, \ldots, j),$$

die alle die gleiche, durch

$$P(X_k = r) = e^{-\lambda} \frac{\lambda^r}{r!} \qquad (r = 0, 1, 2, \ldots; \lambda > 0)$$

bestimmte *Poissonsche Verteilung* haben mögen.

Um die Verteilung der durch (9.8.1) bestimmten Stichprobenfunktion $\bar{X}$ zu finden, beachten wir, daß nach (4.2.6) die charakteristische Funktion der Zufallsvariablen X_k für jedes k durch

$$\varphi(t) = e^{\lambda(e^{it}-1)}$$

bestimmt ist.

Nach (4.2.15) und (4.4.3) hat die charakteristische Funktion der Zufallsvariablen $\bar{X}$ die Gestalt

$$\varphi_1(t) = \exp\left[j\lambda\left(e^{\frac{it}{j}} - 1\right)\right].$$

Dieser Ausdruck stellt die charakteristische Funktion einer Zufallsvariablen mit einer modifizierten Poissonschen Verteilung dar; $\bar{X}$ kann die Werte

$$0, \frac{1}{j}, \frac{2}{j}, \frac{3}{j}, \ldots$$

mit den Wahrscheinlichkeiten

$$P\left(\bar{X} = \frac{l}{j}\right) = \frac{(j\lambda)^l e^{-j\lambda}}{l!} \qquad (l = 0, 1, 2, \ldots)$$

annehmen.

9.9. Die Verteilung der Momente und des Korrelationskoeffizienten in einer Stichprobe aus einer normalen Grundgesamtheit

Wir betrachten die zweidimensionale Zufallsvariable (X, Y), die normal verteilt sei. Ihre Dichte sei durch

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2} - 2\frac{\varrho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\} \tag{9.9.1}$$

gegeben. In dieser Formel sind m_1 und m_2 die Mittelwerte der Zufallsvariablen X und Y, σ_1 und σ_2 sind ihre Standardabweichungen, und ϱ ist der Korrelationskoeffizient der Zufallsvariablen X und Y.

Wir betrachten die Stichprobenfunktionen

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k, \qquad \overline{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k,$$

$$S_1 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2}, \qquad S_2 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(Y_k - \overline{Y})^2}$$

und

$$R = \frac{\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y})}{nS_1S_2};$$

dabei sind (X_k, Y_k), $k = 1, 2, \ldots, n$, unabhängige zweidimensionale Zufallsvariable, deren Dichte durch (9.9.1) bestimmt ist.

Hier sind $\overline{X}$ und $\overline{Y}$ Mittelwerte der Stichprobenfunktion X und Y. Analog sind S_1 und S_2 Standardabweichungen der Stichproben der Variablen X und Y, während R der *Stichprobenkorrelationskoeffizient* der Variablen X und Y ist.

Fisher [2] und Romanowsky [1] haben bewiesen, daß die Dichtefunktion der gemeinsamen Verteilung der Zufallsvariablen $(\overline{X}, \overline{Y}, S_1, S_2, R)$ folgende Form hat (wir geben sie ohne Beweis an):

$$f(\bar{x}, \bar{y}, s_1, s_2, r) = u(\bar{x}, \bar{y})\, v(s_1, s_2, r),$$

wobei

$$u(\bar{x}, \bar{y}) = \frac{n}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left\{-\frac{n}{2(1-\varrho^2)} \cdot \left[\frac{(\bar{x}-m_1)^2}{\sigma_1^2} - 2\frac{\varrho(\bar{x}-m_1)(\bar{y}-m_2)}{\sigma_1\sigma_2} + \frac{(\bar{y}-m_2)^2}{\sigma_2^2}\right]\right\}, \tag{9.9.2}$$

$$v(s_1, s_2, r) = \frac{n^{n-1}}{\pi \Gamma(n-2)} \cdot \frac{s_1^{n-2} s_2^{n-2} (1-r^2)^{\frac{n-4}{2}}}{[\sigma_1^2 \sigma_2^2 (1-\varrho^2)]^{\frac{n-1}{2}}}$$

$$\times \exp\left\{-\frac{n}{2\sigma_1^2\sigma_2^2(1-\varrho^2)}[\sigma_2^2 s_1^2 - 2\varrho r \sigma_1 \sigma_2 s_1 s_2 + \sigma_1^2 s_2^2]\right\} \quad (9.9.3)$$

für $s_1 > 0$, $s_2 > 0$, $r^2 < 1$ und $v(s_1, s_2, r) = 0$ im übrigen Gebiet von (s_1, s_2, r) ist.

Damit hat $(\overline{X}, \overline{Y})$ eine zweidimensionale Normalverteilung mit

$$E(\overline{X}) = m_1, \quad E(\overline{Y}) = m_2, \quad D^2(\overline{X}) = \frac{\sigma_1^2}{n}, \quad D^2(\overline{Y}) = \frac{\sigma_2^2}{n},$$

während der Korrelationskoeffizient der Zufallsvariablen $\overline{X}$ und $\overline{Y}$ gleich ϱ ist.

Aus dem Fisher-Romanowskyschen Satz ergibt sich die äußerst wichtige Folgerung, daß *die Zufallsvariable* $(\overline{X}, \overline{Y})$ *von der Zufallsvariablen* (S_1, S_2, R) *unabhängig ist*. Wir beachten, daß wir aus (9.5.13) für $i = 1, 2$

$$E(S_i^2) = \frac{n-1}{n}\sigma_i^2,$$

$$D^2(S_i^2) = \frac{2(n-1)}{n^2}\sigma_i^4$$

erhalten.

Mit $f(r)$ sei die Dichtefunktion der Zufallsvariablen R bezeichnet. Wie in 3.6.F setzen wir $|M| = \sigma_1^2\sigma_2^2(1-\varrho^2)$ und erhalten aus (9.9.3) die Beziehung

$$f(r) = \frac{n^{n-1}(1-r^2)^{\frac{n-4}{2}}}{\pi\Gamma(n-2)|M|^{\frac{n-1}{2}}}$$

$$\times \int_0^\infty \int_0^\infty s_1^{n-2} s_2^{n-2} \exp\left\{-\frac{n}{2|M|}[\sigma_2^2 s_1^2 - 2\varrho r \sigma_1\sigma_2 s_1 s_2 + \sigma_1^2 s_2^2]\right\} ds_1\, ds_2.$$

Da

$$\exp\left(\frac{n}{|M|}\varrho r \sigma_1 \sigma_2 s_1 s_2\right) = \sum_{j=0}^\infty \frac{1}{j!}\left(\frac{n}{|M|}\varrho r \sigma_1\sigma_2 s_1 s_2\right)^j$$

ist, gilt

$$f(r) = \frac{n^{n-1}(1-r^2)^{\frac{n-4}{2}}}{\pi\Gamma(n-2)|M|^{\frac{n-1}{2}}} \sum_{j=0}^\infty \left[\frac{1}{j!}\left(\frac{n}{|M|}\varrho r\sigma_1\sigma_2\right)^j \int_0^\infty s_1^{n-2+j}\exp\left(-\frac{n}{2|M|}\sigma_2^2 s_1^2\right) ds_1\right.$$

$$\left. \times \int_0^\infty s_2^{n-2+j} \exp\left(-\frac{n}{2|M|}\sigma_1^2 s_2^2\right) ds_2\right].$$

Wenn wir $z_1 = s_1^2$ und $z_2 = s_2^2$ setzen und (5.8.5) benutzen, erhalten wir nach Integration

$$f(r) = \frac{2^{n-3}}{\pi(n-3)!}(1-\varrho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}} \sum_{j=0}^{\infty} \Gamma^2\left(\frac{n+j-1}{2}\right)\frac{(2\varrho r)^j}{j!}.$$

Wir machen von der Formel

$$\Gamma^2(p) = 2^{1-2p}\Gamma(2p)B\left(p, \frac{1}{2}\right)$$

Gebrauch und setzen $p = \dfrac{n+j-1}{2}$, um

$$\sum_{j=0}^{\infty} \Gamma^2\left(\frac{n+j-1}{2}\right)\frac{(2\varrho r)^j}{j!}$$

$$= \frac{(n-2)!}{2^{n-2}} \int_0^1 \sum_{j=0}^{\infty} \binom{n+j-2}{j} (\varrho r)^j z^{\frac{n+j-3}{2}} (1-z)^{-\frac{1}{2}} dz$$

$$= \frac{(n-2)!}{2^{n-3}} \int_0^1 \sum_{j=0}^{\infty} \binom{n+j-2}{j} (\varrho r x)^j x^{n-2} (1-x^2)^{-\frac{1}{2}} dx$$

zu erhalten; dabei substituierten wir $z = x^2$. Da

$$\sum_{j=0}^{\infty} \binom{n+j-2}{j} (\varrho r x)^j = \frac{1}{(1-\varrho r x)^{n-1}}$$

ist, ergibt sich

$$\sum_{j=0}^{\infty} \Gamma^2\left(\frac{n+j-1}{2}\right)\frac{(2\varrho r)^j}{j!} = \frac{(n-2)!}{2^{n-3}} \int_0^1 \frac{x^{n-2}}{(1-\varrho r x)^{n-1}\sqrt{1-x^2}}\, dx.$$

Endgültig erhalten wir also

$$f(r) = \frac{n-2}{\pi}(1-\varrho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}} \int_0^1 \frac{x^{n-2}}{(1-\varrho r x)^{n-1}\sqrt{1-x^2}}\, dx. \tag{9.9.4}$$

Bei festem n hängt die Verteilung des Koeffizienten R nur vom Korrelationskoeffizienten ϱ der Zufallsvariablen X und Y ab.

Verschiedene Werte von ϱ ergeben verschiedene Verteilungen von R. Tafeln für die R-Verteilung bearbeitete DAVID [1].

Wir wollen genauer den Fall $\varrho = 0$ betrachten; für die zweidimensionale Normalverteilung ist dies mit der Voraussetzung der Unabhängigkeit von X und Y gleichwertig. Wir setzen in (9.9.4) $\varrho = 0$ und erhalten

$$f(r) = \frac{n-2}{\pi}(1-r^2)^{\frac{n-4}{2}} \int_0^1 x^{n-2}(1-x^2)^{-\frac{1}{2}}\,dx. \tag{9.9.5}$$

Substituieren wir $z = x^2$, so folgt

$$\begin{aligned} f(r) &= \frac{n-2}{2\pi}(1-r^2)^{\frac{n-4}{2}} \int_0^1 z^{\frac{n-1}{2}-1}(1-z)^{\frac{1}{2}-1}\,dz \\ &= \frac{n-2}{2\pi} B\left(\frac{n-1}{2}, \frac{1}{2}\right)(1-r^2)^{\frac{n-4}{2}} = \frac{1}{B\left(\frac{n-2}{2}, \frac{1}{2}\right)}(1-r^2)^{\frac{n-4}{2}}. \end{aligned} \tag{9.9.6}$$

In (9.9.6) setzen wir

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}, \quad \text{also} \quad 1 - r^2 = \frac{1}{1 + \frac{t^2}{n-2}}. \tag{9.9.7}$$

Dies ergibt

$$\begin{aligned} g(t) &= \frac{1}{B\left(\frac{n-2}{2}, \frac{1}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{n-2}\right)^{\frac{n-4}{2}}} \frac{1}{\sqrt{n-2}} \frac{1}{\left(1 + \frac{t^2}{n-2}\right)^{\frac{3}{2}}} \\ &= \frac{1}{\sqrt{n-2}} \frac{1}{B\left(\frac{n-2}{2}, \frac{1}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{n-2}\right)^{\frac{n-1}{2}}}. \end{aligned} \tag{9.9.8}$$

Vergleichen wir (9.9.8) mit (9.6.5), so sehen wir, daß die Zufallsvariable

$$= \frac{R}{\sqrt{1-R^2}}\sqrt{n-2}$$

im Fall $\varrho = 0$ eine Dichte hat, die mit der Studentschen t-Dichte mit $n-2$ Freiheitsgraden identisch ist.

Beispiel 9.9.1. Die zwei Merkmale X und Y einer Gesamtheit mögen eine zweidimensionale Normalverteilung haben und unabhängig sein. Wir entnehmen dieser Gesamtheit eine einfache Stichprobe vom Umfang $n = 10$. Aus den Stichprobenwerten von X und Y berechnen wir den Korrelationskoeffizienten r und erhalten $r = 0{,}30$. Wie groß ist die Wahrscheinlichkeit dafür, einen nicht kleineren r-Wert als den beobachteten zu erhalten?

Wir wenden die Formel (9.9.7) an und haben

$$P(R \geqq 0{,}3) = P\left(\frac{t}{\sqrt{8 + t^2}} \geqq 0{,}3\right) = P(|t| \geqq 0{,}889).$$

Aus den Tafeln der t-Verteilung mit 8 Freiheitsgraden lesen wir ab, daß die gesuchte Wahrscheinlichkeit 0,40 beträgt. Die Schlüsse, die man hieraus ziehen kann, wollen wir in den nächsten Kapiteln besprechen.

Die Formel (9.9.4) erlaubt uns selbstverständlich auch, die Momente der Zufallsvariablen R zu berechnen. Wir beschränken uns auf die Angabe der Näherungsformeln

$$E(R) \approx \varrho, \qquad D^2(R) \approx \frac{(1 - \varrho^2)^2}{n}. \tag{9.9.9}$$

Die hier angegebenen Näherungsformeln kann man aber nur für genügend große n, d. h. für $n \geqq 500$ benutzen, weil die Verteilung der Zufallsvariablen R sehr asymmetrisch ist. Diese Verteilung strebt zwar für $n \to \infty$ gegen eine Normalverteilung; die Konvergenz geht jedoch sehr langsam vor sich. Beachtenswert ist die von Fisher [3], [7] entdeckte Tatsache, daß die Zufallsvariable

$$U = \frac{1}{2} \log \frac{1 + R}{1 - R}$$

sogar für kleine n eine asymptotische Normalverteilung

$$N\left(\frac{1}{2} \log \frac{1 + \varrho}{1 - \varrho} + \frac{\varrho}{2(n - 1)};\ \sqrt{\frac{1}{n - 3}}\right)$$

hat.

B. Wir betrachten die l-dimensionale $(l \geqq 2)$ Zufallsvariable $(X_1, X_2, \ldots, X_l)$ mit einer Normalverteilung, deren Dichtefunktion durch die Formel (5.11.8) gegeben ist, mit den Erwartungswerten $E(X_j) = m_i$ und mit den Varianzen und Kovarianzen $E(X_j - m_j)(X_i - m_i) = \lambda_{ji}$, wobei die Matrix M der λ_{ji} $(j, i = 1, \ldots, l)$ die Determinante $|M| \neq 0$ hat. Die Zufallsvektoren $(X_{k1}, \ldots, X_{kl})$ mit $k = 1, 2, \ldots, n$ $(n > l)$ seien unabhängig und mögen die gleiche Verteilung haben wie der Vektor $(X_1, X_2, \ldots, X_l)$. Wir betrachten ferner die Mittelwerte der Stichprobenfunktion $\overline{X}_j$, die Varianzen und Kovarianzen der Stichprobenfunktion W_{ji} und die Matrix Q, die durch die Formeln (9.6.11), (9.6.12) bzw. (9.6.13) definiert sind. Die gemeinsame Verteilung des Vektors

$$(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_l, W_{11}, W_{12}, \ldots, W_{l-1,l}, W_{ll})$$

wurde von Wishart [1] gefunden. Er bewies, daß die Vektoren $(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_l)$ und $(W_{11}, W_{12}, \ldots, W_{l-1,l}, W_{ll})$ unabhängig sind, wobei $(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_l)$ eine

l-dimensionale Normalverteilung mit $E(\overline{X}_j) = m_j$ und

$$E\big((\overline{X}_j - m_j)(\overline{X}_i - m_i)\big) = \lambda_{ij}/n$$

hat, während die Dichtefunktion $g(w_{11}, w_{12}, \ldots, w_{l-1,l}, w_{ll})$ des zweiten Vektors in dem Gebiet verschwindet, in dem die Matrix Q nicht positiv definit ist, und im übrigen Gebiet die Form

$$\begin{aligned} g(w_{11}, w_{12}, \ldots, w_{l-1,l}, w_{ll}) \\ = \frac{1}{\pi^{l(l-1)/4}\,\Gamma\left(\frac{1}{2}(n-1)\right)\cdots\Gamma\left(\frac{1}{2}(n-l)\right)} \left(\frac{n^l}{2^l\,|M|}\right)^{(n-1)/2} \\ \times\ |Q|^{(n-l-2)/2} \exp\left(-\frac{n}{2\,|M|}\sum_{i,j=1}^{l} |M_{ji}|\,w_{ji}\right) \end{aligned} \tag{9.9.10}$$

hat, wobei $|M_{ji}|$ das algebraische Komplement des Ausdrucks λ_{ji} in der Determinante $|M|$ ist.

Die durch (9.9.10) gegebene Verteilung heißt die *Wishartsche Verteilung.* Neben dem geometrischen Beweis von Wishart findet man den analytischen Beweis der Formel (9.9.10) bei Hsu [2].

Das Analogon zur verallgemeinerten Varianz für die Grundgesamtheit (siehe 6.3.F) ist die verallgemeinerte Varianz der Stichprobe, und zwar ist sie gleich der Determinante $|Q|$ der Matrix der Varianzen und Kovarianzen der Stichprobe. Mit der Verteilung und den Momenten der Zufallsvariablen $|Q|$ haben sich Wilks [1] und Kullback [1] beschäftigt.

9.10. Die Verteilung der Regressionskoeffizienten

In 3.7 und 3.8 zeigten wir, daß für eine normal verteilte zweidimensionale Zufallsvariable (X, Y) die Regressionskurven erster Art Geraden sind, die mit den Regressionsgeraden zweiter Art identisch sind. Die Koeffizienten dieser Geraden sind durch die Formeln (3.8.4) bzw. (3.8.5) gegeben. Diese Formeln enthalten konstante Koeffizienten. In praktischen, statistischen Fragen sind diese Koeffizienten jedoch Zufallsvariable, und zwar Funktionen der Zufallsvariablen $\overline{X}$, $\overline{Y}$, S_1, S_2 und R. Wir betrachten also die Stichprobenfunktionen

$$A = R\frac{S_2}{S_1}, \quad B = \overline{Y} - R\frac{S_2}{S_1}\overline{X} = \overline{Y} - A\overline{X}. \tag{9.10.1}$$

Ähnlich erhalten wir für die Koeffizienten der zweiten Regressionsgeraden

$$A' = R\frac{S_1}{S_2}, \quad B' = \overline{X} - R\frac{S_1}{S_2}\overline{Y}. \tag{9.10.2}$$

Die Verteilung des Zufallsvektors $(\overline{X}, \overline{Y}, S_1, S_2, R)$ sei durch (9.9.2), (9.9.3) gegeben. Wir bestimmen die Verteilung der Stichprobenfunktion A. Zu diesem Zweck schreiben wir zuerst die Dichte $g(s_1^2, s_2^2, r)$ der Zufallsvariablen (S_1^2, S_2^2, R) hin. Aus (9.9.3) erhalten wir sofort

$$g(s_1^2, s_2^2, r) = \frac{n^{n-1}}{4\pi\Gamma(n-2)\,|M|^{\frac{n-1}{2}}}\, s_1^{n-3} s_2^{n-3} (1-r^2)^{\frac{n-4}{2}} \times \exp\left[-\frac{n}{2|M|}(\sigma_2^2 s_1^2 - 2\varrho r \sigma_1\sigma_2 s_1 s_2 + \sigma_1^2 s_2^2)\right] \tag{9.10.3}$$

mit

$$|M| = \sigma_1^2\sigma_2^2(1-\varrho^2).$$

Wenn wir die Zufallsvariable (S_1^2, S_2^2, A) einführen, ergibt sich als ihre Dichtefunktion

$$f(s_1^2, s_2^2, a) = \frac{n^{n-1}}{4\pi\Gamma(n-2)\,|M|^{\frac{n-1}{2}}}\,(s_2^2 - s_1^2 a^2)^{\frac{n-4}{2}} \cdot \exp\left(-\frac{n}{2|M|}\sigma_1^2 s_2^2\right) s_1^{n-2} \times \exp\left[-\frac{n}{2|M|}(\sigma_2^2 s_1^2 - 2\varrho\,\sigma_1\sigma_2 s_1^2 a)\right]. \tag{9.10.4}$$

Da aus (9.10.1) folgt, daß $s_2^2 - s_1^2 a^2 \geqq 0$ ist, integrieren wir (9.10.4) zuerst nach s_2^2 im Gebiet $s_2^2 - s_1^2 a^2 \geqq 0$ und dann nach s_1^2 im Intervall $(0, \infty)$. Wenden wir die Substitution $y = s_2^2 - s_1^2 a^2$ an, so erhalten wir

$$\begin{aligned}
&\int\limits_{s_2^2 - s_1^2 a^2 \geqq 0} (s_2^2 - s_1^2 a^2)^{\frac{n-4}{2}} \exp\left(-\frac{n}{2|M|}\sigma_1^2 s_2^2\right) ds_2^2 \\
&= \exp\left(-\frac{n}{2|M|}\sigma_1^2 s_1^2 a^2\right) \int\limits_0^\infty y^{\frac{n-4}{2}} \exp\left(-\frac{n\sigma_1^2}{2|M|}y\right) dy \\
&= \frac{\Gamma\left(\frac{n-2}{2}\right) 2^{\frac{n-2}{2}} |M|^{\frac{n-2}{2}}}{n^{\frac{n-2}{2}}\, \sigma_1^{2\,\frac{n-2}{2}}} \exp\left(-\frac{n}{2|M|}\sigma_1^2 s_1^2 a^2\right).
\end{aligned} \tag{9.10.5}$$

Weiter haben wir

$$\int\limits_0^\infty s_1^{2\,\frac{n-2}{2}} \exp\left[-\frac{n}{2|M|}(\sigma_2^2 - 2\varrho\,\sigma_1\sigma_2 a + \sigma_1^2 a^2)\, s_1^2\right] ds_1^2 = \frac{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}} |M|^{\frac{n}{2}}}{n^{\frac{n}{2}}\,(\sigma_2^2 - 2\varrho\,\sigma_1\sigma_2 a + \sigma_1^2 a^2)^{\frac{n}{2}}}. \tag{9.10.6}$$

Aus den Formeln (9.10.4) bis (9.10.6) erhalten wir nach Berücksichtigung der Identität

$$\Gamma(2p) = \frac{2^{2p-1}\Gamma(p)\Gamma\left(p+\frac{1}{2}\right)}{\sqrt{\pi}}$$

die Dichtefunktion $h(a)$ der Stichprobenfunktion A, nämlich

$$h(a) = \frac{\Gamma\left(\frac{n}{2}\right)|M|^{\frac{n-1}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\sigma_1^{\frac{n-2}{2}}(\sigma_2^2 - 2\varrho\sigma_1\sigma_2 a + \sigma_1^2 a^2)^{\frac{n}{2}}}. \tag{9.10.7}$$

Diese Formel gaben K. PEARSON [3] und ROMANOWSKY [2] an. Die Dichtefunktion $h_1(a')$ der Zufallsvariablen A' erhalten wir, indem wir einfach in (9.10.7) die Rollen von σ_1 und σ_2 miteinander vertauschen.

Wir möchten es dem Leser überlassen, die Verteilung der Stichprobenfunktion $B = \overline{Y} - A\overline{X}$ zu finden. Man braucht dazu nur zu beachten, daß $(\overline{X}, \overline{Y})$ eine durch (9.9.2) gegebene zweidimensionale Normalverteilung und A eine durch (9.10.7) gegebene Verteilung hat und daß die Zufallsvariablen $(\overline{X}, \overline{Y})$ und A unabhängig sind.

Wir bestimmen jetzt $E(A)$ und $D^2(A)$.

Zur Abkürzung schreiben wir

$$K = \frac{\Gamma\left(\frac{n}{2}\right)|M|^{\frac{n-1}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\sigma_1^{n-2}},$$

$$V = \sigma_2^2 - 2\varrho\sigma_1\sigma_2 a + \sigma_1^2 a^2.$$

Wir beachten die Formel (9.10.7) und haben für $n > 2$

$$E(A) = K\int_{-\infty}^{\infty}\frac{a}{V^{\frac{n}{2}}}\,da = \frac{K}{2\sigma_1^2}\int_{-\infty}^{\infty}\frac{2\sigma_1^2 a - 2\varrho\sigma_1\sigma_2}{V^{\frac{n}{2}}}\,da + \varrho\frac{\sigma_2}{\sigma_1}K\int_{-\infty}^{\infty}\frac{da}{V^{\frac{n}{2}}} = \varrho\frac{\sigma_2}{\sigma_1}. \tag{9.10.8}$$

Analog ist für $n > 3$

$$E(A^2) = K\int_{-\infty}^{\infty}\frac{a^2}{V^{\frac{n}{2}}}\,da = \frac{1}{\sigma_1^2}K\int_{-\infty}^{\infty}\frac{1}{V^{\frac{n-2}{2}}}\,da + 2\varrho\frac{\sigma_2}{\sigma_1}K\int_{-\infty}^{\infty}\frac{a\,da}{V^{\frac{n}{2}}} - \frac{\sigma_2^2}{\sigma_1^2}K\int_{-\infty}^{\infty}\frac{1}{V^{\frac{n}{2}}}\,da.$$

Wenn wir noch berücksichtigen, daß aus (9.10.7) (wir setzen nur $n-2$ statt n)

$$\int_{-\infty}^{\infty} \frac{1}{V^{\frac{n-2}{2}}}\, da = \frac{\sqrt{\pi}\,\Gamma\left(\frac{n-3}{2}\right) \sigma_1^{n-4}}{\Gamma\left(\frac{n-2}{2}\right) |M|^{\frac{n-3}{2}}}$$

folgt, so ergibt sich nach einfachen Rechnungen

$$E(A^2) = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{1 - 4\varrho^2 + \varrho^2 n}{n-3}.$$

Daraus und aus (9.10.8) erhalten wir

$$D^2(A) = \frac{1}{n-3} \frac{\sigma_2^2}{\sigma_1^2} (1 - \varrho^2). \tag{9.10.9}$$

Mittelwert und Dispersion der Zufallsvariablen

$$B = \overline{Y} - A\overline{X}$$

sind leicht zu berechnen, da wir schon die entsprechenden Momente der Zufallsvariablen $\overline{Y}$, A und $\overline{X}$ kennen und außerdem wissen, daß $(\overline{X}, \overline{Y})$ und A unabhängig sind. Wir erhalten

$$E(B) = m_2 - \varrho \frac{\sigma_2}{\sigma_1} m_1, \tag{9.10.10}$$

$$D^2(B) = \frac{\sigma_2^2}{n} \frac{n-2-\varrho^2}{n-3} + \frac{\sigma_2^2}{\sigma_1^2} \frac{m_1^2(1-\varrho^2)}{n-3}. \tag{9.10.11}$$

Vernachlässigen wir in der letzten Formel die Glieder der Größenordnung unter n^{-1}, so ergibt sich

$$D^2(B) \approx \frac{\sigma_2^2}{(n-3)\sigma_1^2} [\sigma_1^2 + m_1^2(1-\varrho^2)]. \tag{9.10.12}$$

Sehr vorteilhaft ist die Substitution

$$t = \frac{\sigma_1 \sqrt{n-1}}{\sigma_2 \sqrt{1-\varrho^2}} (A - \alpha), \tag{9.10.13}$$

wobei $\alpha = \varrho \frac{\sigma_2}{\sigma_1}$ ist. Es gilt dann

$$1 + \frac{t^2}{n-1} = \frac{A^2\sigma_1^2 - 2\varrho\sigma_1\sigma_2 A + \sigma_2^2}{|M|} \sigma_1^2.$$

Aus (9.10.7) erhalten wir damit für die Dichtefunktion $f(t)$ der Zufallsvariablen t den Ausdruck

$$f(t) = \frac{1}{\sqrt{n-1}} \frac{1}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{\frac{n}{2}}}. \tag{9.10.14}$$

Folglich hat die durch (9.10.13) bestimmte Zufallsvariable t eine Studentsche t-Verteilung mit $n - 1$ Freiheitsgraden.

In der Formel (9.10.13) bestimmten wir die Zufallsvariable t als Funktion der Zufallsvariablen A und der Parameter σ_1, σ_2 und ϱ der Grundgesamtheit. In der Praxis kennen wir aber meistens die Werte dieser Parameter nicht. Von großer Bedeutung ist deshalb die von Bartlett [1] festgestellte Tatsache, daß die durch die Formel

$$t_1 = \frac{S_1 \sqrt{n-2}}{S_2 \sqrt{1-R^2}} (A - \alpha) \tag{9.10.15}$$

definierte Zufallsvariable t_1 eine Studentsche t-Verteilung mit $n - 2$ Freiheitsgraden hat. Der Leser beachte, daß man aus (9.10.13) die Formel (9.10.15) erhält, wenn man die Parameterwerte σ_1, σ_2 und ϱ durch S_1, S_2 bzw. R sowie $n - 1$ durch $n - 2$ und t durch t_1 ersetzt.

Den Beweis des Bartlettschen Satzes bringen wir hier nicht, da er leicht aus (9.9.3) hergeleitet werden kann. Dazu hat man nur die gemeinsame Verteilung der S_1, S_2 und t_1 zu bestimmen und dann die Randverteilung von t_1 zu berechnen.

9.11. Die Grenzverteilungen von Stichprobenmomenten

In den vorherigen Paragraphen besprachen wir exakte Verteilungen von Stichprobenfunktionen. Diese Verteilungen zu finden, ist meistens sehr schwierig, und oft erhält man Formeln, die nur wenig praktischen Wert haben. Dagegen kann man ziemlich leicht die Grenzverteilungen einer umfangreichen Klasse von Stichprobenfunktionen angeben, die häufig in der Praxis vorkommen. Besitzt die Stichprobe einen genügend großen Umfang n, so nähert sich die Verteilung der untersuchten Stichprobenfunktion der Grenzverteilung, die man für $n \to \infty$ erhält. Hierauf beruht die Anwendbarkeit der Grenzverteilungen von Stichprobenfunktionen.

Die Ausführungen dieses Paragraphen finden ausschließlich bei genügend umfangreichen Stichproben Anwendung. Es gibt keine für alle Stichprobenfunktionen allgemein gültige Vorschrift, die vorschreiben würde, für welche n man eine Stichprobe als „groß" ansehen kann. Wann eine Stichprobe „groß" ist, hängt davon ab, wie gut die Verteilung der untersuchten Stichprobenfunktion gegen die Grenzverteilung strebt.

Es seien $X_1, X_2, \ldots, X_n$ unabhängige Zufallsvariable mit derselben Verteilung wie die Zufallsvariable X in der Grundgesamtheit. Wir nehmen an, daß das k-te Moment $m_k = E(X^k)$ existiert, und untersuchen die Stichprobenfunktion

$$A_k = \frac{1}{n} \sum_{r=1}^{n} X_r^k.$$

Wir sehen, daß

$$E(A_k) = \frac{1}{n} \sum_{r=1}^{n} E(X_r^k)$$

gilt. Da nach Voraussetzung $E(X_r^k) = E(X^k)$ für $r = 1, 2, \ldots, n$ ist, haben wir

$$E(A_k) = \frac{1}{n} \sum_{r=1}^{n} E(X^k) = m_k.$$

Die Stichprobenfunktion A_k nennen wir das *k-te Stichprobenmoment*. Aus dem Chintschinschen Satz 6.11.4 folgt, daß für jedes k die Folge $\{A_k\}$ der Stichprobenmomente stochastisch gegen das Moment m_k der Grundgesamtheit konvergiert, falls n gegen ∞ strebt. Aus Satz 6.14.2 von Slutski folgt weiter, daß die Folge $\{B_k\}$ der zentralen Stichprobenmomente, die durch

$$B_k = \frac{1}{n} \sum_{r=1}^{n} (X_r - A_1)^k$$

definiert sind, für $n \to \infty$ gegen das zentrale Moment μ_k der Grundgesamtheit konvergiert. Somit konvergiert die Stichprobendispersion für $n \to \infty$ stochastisch gegen die Dispersion der Gesamtheit.

Setzen wir die Existenz des $2k$-ten Moments m_{2k} der Zufallsvariablen X voraus, so können wir eine Formel für die Dispersion der Zufallsvariablen A_k herleiten. Für $r = 1, 2, \ldots, n$ gilt nämlich die Beziehung

$$D^2(X_r^k) = D^2(X^k) = E(X^{2k}) - [E(X^k)]^2 = m_{2k} - m_k^2$$

und wegen der Unabhängigkeit der Zufallsvariablen $X_1, X_2, \ldots, X_n$ die Formel

$$D^2(A_k) = \frac{1}{n^2} \sum_{r=1}^{n} D^2(X_r^k) = \frac{n}{n^2} D^2(X^k) = \frac{m_{2k} - m_k^2}{n}.$$

Aus dem Lindeberg-Lévyschen Grenzwertsatz 6.8.1 folgt, daß für $n \to \infty$ die Folge der Verteilungsfunktionen der normierten Zufallsvariablen

$$Y_n = \frac{\sum_{r=1}^{n} X_r^k - n m_k}{\sqrt{n D^2(X^k)}} = \frac{A_k - m_k}{\sqrt{m_{2k} - m^2}} \sqrt{n}$$

gegen die Grenzverteilungsfunktion $\Phi(x)$ strebt. Wir können also sagen: Existiert das endliche Moment m_{2k} in der Grundgesamtheit, dann ist das Moment A_k der Ordnung k asymptotisch normal

$$N\left(m_k;\ \sqrt{\frac{m_{2k}-m_k^2}{n}}\right)$$

verteilt.

Beispiel 9.11.1. Eine Gruppe von Karteikarten entspreche den in einer Versicherungsgesellschaft Versicherten. Die Versicherten teilen wir in zwei Kategorien ein: Zur ersten sollen diejenigen Versicherten gehören, die eine Familie zu erhalten haben, zur zweiten alleinstehende Personen. Die erste Gruppe sei auf der Karteikarte mit einer 1, die zweite mit einer 0 bezeichnet. Der Anteil der Versicherten, die zur ersten Kategorie gehören, betrage p. Der Anteil der Versicherten, die zur zweiten Kategorie gehören, beträgt also $q = 1 - p$.

Dieser Kartengruppe entnehmen wir eine einfache Stichprobe vom Umfang n. Dabei werde vor dem Herausziehen einer jeden neuen Karte die schon gezogene Karte zurückgelegt.

Wir bezeichnen das Ergebnis der r-ten Ziehung mit x_r; x_r kann entweder den Wert 1 oder den Wert 0 annehmen und ist der beobachtete Wert der Zufallsvariablen X_r, die eine Null-Eins-Verteilung hat. Das arithmetische Mittel $\bar{x}$ der Werte x_r stellt den beobachteten Wert der Zufallsvariablen

$$A_1 = \bar{X} = \frac{1}{n}\sum_{r=1}^{n} X_r$$

dar. Es sei nun n eine große Zahl, zum Beispiel $n = 100$, und es sei $p = 0{,}3$. Wie groß ist die Wahrscheinlichkeit dafür, daß wir einen $\bar{x}$-Wert erhalten, dessen Absolutwert sich von $p = 0{,}30$ höchstens um 0,02 unterscheidet?

Wir benutzen den eben bewiesenen Satz, der aussagt, daß das Stichprobenmoment A_k asymptotisch normal

$$N\left(m_k;\ \sqrt{\frac{m_{2k}-m_k^2}{n}}\right)$$

verteilt ist, wenn das Moment m_{2k} existiert.

Für unser Beispiel ist

$$m_1 = m_2 = p = 0{,}3 .$$

Also ist die Stichprobenfunktion A_1 asymptotisch normal $N(0{,}3;\, 0{,}0458)$ verteilt. Wir suchen die Wahrscheinlichkeit

$$P(|A_1 - 0{,}3| \geqq 0{,}02) = P\left(\frac{|A_1 - 0{,}3|}{0{,}0458} \geqq 0{,}44\right).$$

Aus den Tafeln der Normalverteilung finden wir, daß diese Wahrscheinlichkeit gleich 0,66 ist.

9.12. Aufgaben und Ergänzungen

1. a) Man zeige, daß die Dichtefunktion $g(t)$ der durch (9.6.5) definierten Studentschen t-Verteilung die Beziehung

$$\lim_{n\to\infty} g(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$$

erfüllt.

b) Nach Gewinnung dieses Ergebnisses zeige man, daß auch die Beziehung (9.6.6) gilt.

2. Man zeige, daß die Stichprobenfunktionen $\bar{X}$ und S^2, d. h. der Mittelwert und die Varianz von einfachen aus einer Gesamtheit entnommenen Stichproben, in der das Merkmal X eine symmetrische Verteilung hat, nicht korreliert sind.

3. Man leite die Formel (9.6.7) in der gleichen Weise her wie die Dichtefunktion der Studentschen t-Verteilung.

4. Man leite die Formel für die Dichtefunktion der Zufallsvariablen $Y = t^2$ her, wobei t die durch (9.6.5) gegebene Verteilung hat, und vergleiche das Ergebnis mit (9.6.15) für $r = 1$.

5. In der Formel (9.6.14) für die Hotellingsche T^2-Verteilung setze man $r = 2$ und $n = 4$. Sodann bestimme man y-Werte derart, daß a) $P(T^2 > y) = 0{,}05$ und b) $P(T^2 > y) = 0.01$ ist.

6. Man leite die Dichtefunktion der Snedecorschen Stichprobenfunktion F her.

7. Es seien X_k $(k = 1, 2, \ldots, n_1)$ und Y_l $(l = 1, 2, \ldots, n_2)$ unabhängig, und es möge X_k die Verteilung $N(0; \sigma_1)$ und Y_l die Verteilung $N(0; \sigma_2)$ haben. Ferner sei

$$Z = \frac{1}{2} \log \frac{n_1(n_2 - 1)\, S_1^2}{n_2(n_1 - 1)\, S_2^2}.$$

Man zeige, daß die Dichtefunktion $h(z)$ der Zufallsvariablen Z die Form

$$h(z) = \frac{2\sigma_1^{-r_1}\sigma_2^{-r_2} r_1^{r_1/2} r_2^{r_2/2}}{B\left(\frac{r_1}{2}, \frac{r_2}{2}\right)} \frac{e^{r_1 z}}{\left(\frac{r_1 e^{2z}}{\sigma_1^2} + \frac{r_2}{\sigma_2^2}\right)^{(r_1+r_2)/2}}$$

mit $r_i = n_i - 1$ $(i = 1, 2)$ hat.

8. Man bestimme die Verteilung der Stichprobenfunktion $\bar{X}$ einer einfachen n-gliedrigen Stichprobe aus einer Gesamtheit, in der X der Gammaverteilung genügt.

9. Es sei $\bar{X}$ der Mittelwert einer einfachen n-gliedrigen Stichprobe aus einer Gesamtheit, in der das Merkmal X die durch (5.6.6) gegebene Gleichverteilung aufweist. Man zeige, daß die Dichtefunktion $f(t)$ der Zufallsvariablen $\bar{X}$ die Form

$$g(x) = \frac{n^n}{(n-1)!} \sum_{k=0}^{j} (-1)^k \binom{n}{k} \left(x - \frac{k}{n}\right)^{n-1} \qquad \left(\frac{j}{n} \leqq x \leqq \frac{j+1}{n}\right)$$

hat, wobei $j = 0, 1, \ldots, n-1$ ist.

10. Es sei $(X_1, X_2, \ldots, X_l)$, $l \geqq 3$, der Vektor einer Grundgesamtheit, und es habe die Matrix M der Momente zweiter Ordnung λ_{ji} $(j, i = 1, 2, \ldots, l)$ die Determinante $|M| \neq 0$. Die $(X_{k1}, \ldots, X_{kl})$ $(k = 1, \ldots, n)$ seien unabhängige Beobachtungswerte des Vektors $(X_1, X_2, \ldots, X_l)$. Ferner sei Q die Matrix der Stichprobenmomente zweiter Ordnung,

d. h. die Matrix der Elemente W_{ji}, die durch (9.6.12) und (9.6.13) definiert sind. R_{ji} $(j, i = 1, \ldots, l)$ sei der Stichprobenkorrelationskoeffizient der Zufallsvariablen X_i und X_j, und es sei C die Matrix mit den Elementen R_{ji} und $|C|$ die zugehörige Determinante (wobei $R_{jj} = 1$ angenommen wird).
Man zeige, daß der Vektor $(R_{12}, R_{13}, \ldots, R_{l-1,l})$ die durch die Formel

$$\frac{\left[\Gamma\left(\frac{1}{2}(n-1)\right)\right]^{l-1}}{\pi^{l(l-1)/4}\,\Gamma\left(\frac{1}{2}(n-2)\right)\cdots\Gamma\left(\frac{1}{2}(n-l)\right)}\,|C|^{(n-1-2)/2}$$

gegebene Dichtefunktion hat, wenn $(X_1, X_2, \ldots, X_l)$ eine Normalverteilung mit $\lambda_{jj} = 1$ und $\lambda_{ji} = 0$ $(j \neq i)$ besitzt.

11. (Fortsetzung.) Der *partielle Stichprobenkorrelationskoeffizient* von X_1 und X_2 bezüglich $X_3, \ldots, X_l$ wird durch die Formel (vgl. Aufgabe 3.9.15)

$$R_{12\cdot3\ldots l} = -\frac{|Q_{12}|}{\sqrt{|Q_{11}|\cdot|Q_{22}|}} = -\frac{|C_{12}|}{\sqrt{|C_{11}|\cdot|C_{22}|}}$$

definiert, wobei $|Q_{ji}|$ bzw. $|C_{ji}|$ die algebraischen Komplemente des Ausdrucks W_{ji} der Determinante $|Q|$ bzw. des Ausdrucks R_{ji} der Determinante $|C|$ bedeutet.

a) Man beweise: Hat $(X_1, X_2, \ldots, X_l)$ die Normalverteilung, so hat die Dichtefunktion $f(r_{12\cdot3\ldots l})$ der Zufallsvariablen $R_{12\cdot3\ldots l}$ die Form (FISHER [13])

$$f(r_{12\cdot3\ldots l}) = \frac{n-l}{\pi}\,(1-\varrho^2_{12\cdot3\ldots l})^{(n-l+1)/2}(1-r^2_{12\cdot3\ldots l})^{(n-1-2)/2}$$
$$\times\int_0^1 \frac{x^{n-1}}{(1-\varrho_{12\cdot3\ldots l}\,r_{12\cdot3\ldots l})^{n-l+1}\sqrt{1-x^2}}\,dx.$$

b) Man zeige: In dem Spezialfall, daß $(X_1, X_2, \ldots, X_l)$ eine Normalverteilung hat und $\varrho_{12\cdot3\ldots l} = 0$ ist, gilt

$$f(r_{12\cdot3\ldots l}) = \frac{1}{B\left(\frac{1}{2}(n-l), \frac{1}{2}\right)}\,(1-r^2_{12\cdot3\ldots l})^{(n-l-2)/2}.$$

c) Man vergleiche diese Formeln mit (9.9.4) und (9.9.6). Wodurch ist n in den Formeln (9.9.4) und (9.9.6) zu ersetzen, um die hier angegebenen Formeln zu erhalten?

12. (Fortsetzung.) Man zeige, daß die Zufallsvariable

$$\frac{R_{12\cdot3\ldots l}}{\sqrt{1-R^2_{12\cdot3\ldots l}}}\sqrt{n-l}$$

für $\varrho_{12\cdot3\ldots l} = 0$ die Studentsche t-Verteilung mit $n-l$ Freiheitsgraden hat.

13. (Fortsetzung.) Der *mehrfache Korrelationskoeffizient* (siehe Aufgabe 3.9.17) der Zufallsvariablen X_1 bezüglich der Zufallsvariablen $X_2, X_3, \ldots, X_l$ wird durch die Formel

$$R_{1(2\ldots l)} = \sqrt{1-\frac{|Q|}{W_{11}|Q_{11}|}} = \sqrt{1-\frac{|C|}{|C_{11}|}}$$

definiert. Man beweise: Hat $(X_1, X_2, \ldots, X_l)$ eine Normalverteilung und ist $\varrho_{ji} = 0$ $(j, i = 1, \ldots, l;\ j \neq i)$, so hat $R^2_{1(2\ldots l)}$ die Betaverteilung mit den Parametern $\frac{1}{2}(l-1)$, $\frac{1}{2}(n-l)$ (FISHER [14]; die Verteilung von $R^2_{1(2\ldots l)}$ ohne die Voraussetzung $\varrho_{ji} = 0$ hat FISHER [15] angegeben).

14. (Fortsetzung.) Man zeige, daß $nR^2_{1(2\ldots l)}$ für $n \to \infty$ eine χ^2-Verteilung mit $l-1$ Freiheitsgraden hat und somit nicht asymptotisch normal verteilt ist.
In den nachfolgenden Aufgaben 15 bis 21 verwenden wir die Bezeichnungen von 9.11 und setzen voraus, daß sämtliche Momente der Grundgesamtheit existieren, die in diesen Aufgaben auftreten. Die Bedeutung des Symbols $O(\ )$ ist in Aufgabe 8.13.31 erklärt.

15. Man zeige, daß die folgenden Gleichungen gelten:
$$E[(A_1 - m_1)^3] = \frac{\mu_3}{n^2}, \quad E[(A_1 - m_1)^4] = \frac{3\mu_2}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}.$$

16. Man zeige, daß für $j = 1, 2, \ldots$ und $n \to \infty$
$$E[(A_n - m_n)^{2j-1}] = O\left(\frac{1}{n^j}\right), \quad E[(A_n - m_n)^{2j}] = O\left(\frac{1}{n^j}\right)$$
ist.

17. a) Man beweise:
$$E(B_2^2) = \mu_2^2 + \frac{\mu_4 + 3\mu_2^2}{n} - \frac{2\mu_2 - 5\mu_2^2}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}.$$
b) Unter Verwendung dieser Formel und der Beziehung (13.3.4) leite man eine Formel für $D^2(B_2)$ her.
c) Für den Fall, daß das Merkmal X in einer Grundgesamtheit die Normalverteilung $N(m; \sigma)$ hat, zeige man, indem man darüber hinaus die Formel (5.7.7) hinzuzieht, daß
$$D^2(B_2) = \frac{2(n-1)}{n^2}\sigma^4$$
gilt.

18. Man beweise:
$$E\left[(A_1 - m_1)\left(B_2 \frac{n-1}{n}\mu_2\right)\right] = \frac{n-1}{n^2}\mu_3.$$

19. Man beweise:
$$E(B_k) = \mu_k + O\left(\frac{1}{n}\right), \quad D^2(B_k) = \frac{\mu_{2k} - 2k\mu_{k-1}\mu_{k+1} - \mu_k^2 + k^2\mu_2\mu_{k-1}^2}{n} + O\left(\frac{1}{n^2}\right).$$

20. Man zeige, daß das zentrale Moment der Ordnung k einer einfachen Stichprobe für $n \to \infty$ die asymptotische Normalverteilung
$$N\left(\mu_k;\ \sqrt{\frac{\mu_{2k} - 2k\mu_{k-1}\mu_{k+1} - \mu_k^2 + k^2\mu_2\mu_{k-1}^2}{n}}\right)$$
hat.

21. Man bestimme die Grenzverteilung für $n \to \infty$ des Zufallsvektors
$$[\sqrt{n}(A_1 - m_1),\ \sqrt{n}\,(A_2 - m_2)].$$
Hinweis. Man verwende den Satz 6.13.2.

10. DIE VERTEILUNG DER POSITIONSSTICHPROBEN-FUNKTIONEN

10.1. Einleitende Bemerkungen

Im vorigen Kapitel beschäftigten wir uns mit der Verteilung der Stichprobenmomente und mit einigen ihrer Funktionen. Ein Stichprobenmoment ist das Gegenstück zum entsprechenden Moment der Gesamtheit. In diesem Kapitel wollen wir uns mit der Verteilung von Stichprobenfunktionen befassen, die wir *Positionsstichprobenfunktionen* nennen wollen; diese entsprechen den Lageparametern der Gesamtheit. Wir führen hier den Begriff der Stichprobenmediane und allgemeiner den eines Stichprobenquantils ein. Die Stichprobenquantile und ihre Funktionen sind Positionsstichprobenfunktionen. Die Theorie der Verteilungen dieser Stichprobenfunktionen und besonders die Theorie ihrer Grenzverteilungen spielt eine immer größere Rolle in der Wahrscheinlichkeitsrechnung und der mathematischen Statistik.

Die genauen Beziehungen, die zwischen einer Stichprobe und der Grundgesamtheit bestehen, werden vom Satz von GLIWENKO (vgl. 10.10) und den Sätzen von KOLMOGOROFF und SMIRNOW (vgl. 10.11) erfaßt. Diese Sätze gehören zu den grundlegenden Ergebnissen der Theorie der Verteilungen von Positionsstichprobenfunktionen.

10.2. Die Positionsstichprobenfunktionen

In Kapitel 3 beschäftigten wir uns mit den Lageparametern. Lageparameter sind insbesondere die Mediane, die Quantile und das kleinste Element der Stichprobe. In Stichproben von bestimmtem Umfang n wollen wir Zufallsvariable betrachten, die ein Gegenstück zu den Lageparametern bilden. Diese Zufallsvariablen wollen wir *Positionsstichprobenfunktionen* oder kürzer *Positionsfunktionen* nennen.

Diesen Begriff wollen wir nun genau formulieren. Es sei $(X_1, X_2, \ldots, X_n)$ ein Zufallsvektor.

Wir definieren die Zufallsvariable $\zeta_k^{(n)}$, die eine Funktion des Zufallsvektors $X = (X_1, \ldots, X_n)$ sein soll, auf folgende Weise: Jedes Wertetupel $(x_1, x_2, \ldots, x_n)$ dieses Zufallsvektors ordnen wir der Größe nach an und erhalten so eine Folge $x_{r_1}, x_{r_2}, \ldots, x_{r_n}$, die die Ungleichungen

$$x_{r_1} \leqq x_{r_2} \leqq \cdots \leqq x_{r_n}$$

erfüllt. Sind zwei Komponenten x_i und x_j gleich, so ist ihre Reihenfolge gleichgültig.

Definition 10.2.1. Als *Positionsfunktion* $\zeta_k^{(n)}$ bezeichnen wir die Zufallsvariable, die für jede der möglichen Wertegruppen $(x_1, x_2, \ldots, x_n)$ des Vektors $(X_1, \ldots, X_n)$ den k-ten Wert x_{r_k} des entsprechenden, der Größe nach geordneten n-Tupels $(x_{r_1}, x_{r_2}, \ldots, x_{r_n})$ annimmt. Die Zahl k wird der *Rang* von $\zeta_k^{(n)}$ genannt.

Für gegebenes n können wir n derartige Positionsfunktionen bilden, nämlich die Stichprobenfunktionen

$$\zeta_1^{(n)}, \zeta_2^{(n)}, \ldots, \zeta_n^{(n)}.$$

Beispiel 10.2.1. Jede der diskreten Zufallsvariablen X_1, X_2, X_3 möge die Werte 0, 1, 2 annehmen können. Die Zufallsvariable (X_1, X_2, X_3) kann dann die folgenden Wertetupel annehmen:

$$\begin{array}{lllllllll}
(0,0,0), & (0,0,1), & (0,1,0), & (1,0,0), & (0,0,2), & (0,2,0), & (2,0,0), & (1,1,0), & (1,0,1), \\
(0,1,1), & (0,1,2), & (0,2,1), & (2,0,1), & (2,1,0), & (1,0,2), & (1,2,0), & (0,2,2), & (2,0,2), \\
(2,2,0), & (1,1,1), & (1,1,2), & (1,2,1), & (2,1,1), & (1,2,2), & (2,1,2), & (2,2,1), & (2,2,2).
\end{array}$$

Hier können wir drei Positionsfunktionen

$$\zeta_1^{(3)}, \zeta_2^{(3)}, \zeta_3^{(3)}$$

bilden. Die erste nimmt in jedem dieser Tupel den kleinsten Wert an. Entsprechend verhalten sich die beiden übrigen Positionsfunktionen.

Definition 10.2.2. Das Verhältnis $\frac{k}{n}$ bezeichnen wir als den *relativen Rang* der Positionsfunktionen $\zeta_k^{(n)}$.

Um die Grenzverteilung der Positionsfunktionen für $n \to \infty$ zu untersuchen, betrachten wir die Folgen $\{\zeta_k^{(n)}\}$. Dabei kann k eine Funktion der Veränderlichen n sein.

Wir unterscheiden zwei Typen von Positionsfunktionenfolgen.

Definition 10.2.3. Ist $\lim\limits_{n\to\infty} \frac{k}{n} = 0$ oder $\lim\limits_{n\to\infty} \frac{k}{n} = 1$, so bezeichnen wir die Folge $\{\zeta_k^{(n)}\}$ als *äußere* Positionsfunktionenfolge. Ist dagegen $\lim\limits_{n\to\infty} \frac{k}{n} = \lambda$ mit $0 < \lambda < 1$, so bezeichnen wir die Folge $\{\zeta_k^{(n)}\}$ als *zentrale* Positionsfunktionenfolge.

Äußere Positionsfunktionenfolgen erhalten wir, falls wir für k irgendeinen festen Wert, zum Beispiel 1, 2, ... oder eine Zahl von der Form $n - l$, wobei l konstant ist, nehmen. Für $n \to \infty$ haben wir im ersten Fall $\frac{k}{n} \to 0$ und im zweiten $\frac{k}{n} \to 1$.

Beispiel 10.2.2. Einer Grundgesamtheit entnehmen wir eine Stichprobenserie mit wachsendem Umfang n. In jeder dieser Proben beobachten wir den kleinsten Wert des Merkmals X oder, anders ausgedrückt, den Wert der Zufallsvariablen $\zeta_1^{(n)}$. Diese Werte bilden eine äußere Positionsfunktionenfolge.

Äußere Positionsfunktionenfolgen erhält man nicht nur für konstante k oder durch Subtraktion einer konstanten Zahl l von n, sondern zum Beispiel auch dann, wenn k (das im allgemeinen von n abhängig ist) langsamer als n gegen Unendlich wächst. In diesem Fall gilt nämlich $\frac{k}{n} \to 0$ oder $\frac{n-k}{n} \to 1$.

Beispiel 10.2.3. Einer Grundgesamtheit entnehmen wir eine Stichprobenserie mit wachsendem Umfang n. In jeder Stichprobe beobachten wir den Wert der Zufallsvariablen $\zeta_k^{(n)}$ mit $k = [\sqrt{n}]$.[1] Für Stichproben vom Umfang

$$n = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, \ldots$$

haben wir also

$$k = 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, \ldots$$

Die so bestimmte Positionsfunktionenfolge $\{\zeta_k^{(n)}\}$ gehört ebenfalls zu den äußeren, denn es ist

$$\lim_{n\to\infty} \frac{k}{n} = \lim_{n\to\infty} \frac{[\sqrt{n}]}{n} = \lim_{n\to\infty} \frac{\sqrt{n}}{n} = \lim_{n\to\infty} \frac{1}{\sqrt{n}} = 0.$$

Wir wählen nun $k = \left[\frac{n}{2}\right] + 1$, d. h., für ungerade n wählen wir das mittlere Element (der Größe nach gezählt), während wir für gerade $n = 2l$ das $(l+1)$-te Element nehmen. Diese Positionsfunktion $\zeta_k^{(n)}$ nennen wir die *Stichprobenmediane*.

Wenn $k = [n\lambda] + 1$ gilt, wobei λ eine beliebige reelle Zahl ist, $0 < \lambda < 1$, so nennen wir die Positionsfunktion $\zeta_k^{(n)}$ ein *Stichprobenquantil*. Offenbar strebt $\frac{k}{n}$ für $n \to \infty$ gegen λ. Daher stellt die Folge der Stichprobenquantile eine zentrale Positionsfunktionenfolge dar.

Ist k konstant oder gilt $k = n - l + 1$, wobei l konstant ist, so nennen wir $\zeta_k^{(n)}$ *sukzessives Element*.

Funktionen der Positionsfunktionen $\zeta_k^{(n)}$ werden wir gleichfalls als *Positionsstichprobenfunktionen* oder kürzer als *Positionsfunktionen* bezeichnen.

10.3. Die empirische Verteilungsfunktion

A. Es sei $(X_1, X_2, \ldots, X_n)$ ein Zufallsvektor, und es mögen $(\zeta_1^{(n)}, \zeta_2^{(n)}, \ldots, \zeta_n^{(n)})$ sowie $(x_{r_1}, x_{r_2}, \ldots, x_{r_n})$ die gleiche Bedeutung haben wie in 10.2.

Definition 10.3.1. Die Funktion der Variablen x $(-\infty < x < \infty)$, die gleich 0 für $x \leq x_{r_1}$ und gleich $\frac{m}{n}$ $(m = 1, 2, 3, \ldots, n)$ für $x > x_{r_1}$ ist, wobei m den

[1] Unter $[A]$ verstehen wir die größte ganze Zahl, die nicht größer als A ist.

größten Index bedeutet, für den die Ungleichung

$$x_{r_m} < x \tag{10.3.1}$$

erfüllt ist, heißt *empirische Verteilungsfunktion.*

Der allgemeinen Gepflogenheit zufolge werden wir die empirische Verteilungsfunktion mit $S_n(x)$ bezeichnen.

Somit ist $nS_n(x)$ die Anzahl der Stichprobenelemente, die kleiner sind als x. Die Funktion $S_n(x)$ kann also Werte zwischen 0 und 1 annehmen und ist offenbar eine nicht abnehmende Funktion von x. Desgleichen verifiziert man leicht, daß $S_n(x)$ eine linksseitig stetige Funktion von x ist. $S_n(x)$ hat als Funktion von x die Eigenschaften einer Verteilungsfunktion, und wir bezeichnen sie daher als empirische Verteilungsfunktion. Wir weisen jedoch darauf hin, daß $S_n(x)$ für jeden Wert von x eine Zufallsvariable ist.

B. Nun mögen die Zufallsvariablen $X_r (r = 1, 2, \ldots, n)$ unabhängig sein und die gleiche stetige Verteilungsfunktion $F(x)$ haben.

Die $x_{r_1}, x_{r_2}, \ldots, x_{r_n}$ mögen ein der Größe nach geordnetes System von Beobachtungswerten der Zufallsvariablen $X_1, X_2, \ldots, X_n$ bedeuten. Aus der Annahme, daß die Verteilungsfunktion $F(x)$ stetig ist und die Variablen $X_1, X_2, \ldots, X_n$ unabhängig sind, geht hervor, daß die Wahrscheinlichkeit für das Auftreten zweier einander gleicher Werte X_{r_k} und $X_{r_{k+1}}$ gleich 0 ist. Wir können daher annehmen, daß $x_{r_1} < x_{r_2} < \cdots < x_{r_n}$ gilt. Nunmehr können wir die empirische Verteilungsfunktion $S_n(x)$ in der folgenden Form darstellen:

$$S_n(x) = \begin{cases} 0 & \text{für} \quad x \leqq x_{r_1}, \\ \dfrac{k}{n} & \text{für} \quad x_{r_k} < x \leqq x_{r_{k+1}} \quad (k = 1, 2, \ldots, n-1), \\ 1 & \text{für} \quad x > x_{r_n}. \end{cases} \tag{10.3.2}$$

C. Nun seien die $X_1, X_2, \ldots, X_n$ unabhängig und mögen die gleiche Verteilungsfunktion $F(x)$ haben. Aus den üblichen Voraussetzungen folgt, daß für jedes feste x die Beziehung

$$P(X_r < x) = F(x) = p = \text{const} \qquad (r = 1, 2, \ldots, n) \tag{10.3.3}$$

gilt. Somit ist $S_n(x)$ für einen festen Wert von x die Häufigkeit in einem Bernoullischen Versuchsschema und demnach die Realisierungshäufigkeit eines zufälligen Ereignisses in n unabhängigen Versuchen, wobei die Wahrscheinlichkeit für die Realisierung eines Ereignisses für jeden einzelnen Versuch gleich $p = F(x)$ ist. Hieraus erhalten wir (siehe 5.2) für $m = 0, 1, \ldots, n$

$$P\left(S_n(x) = \frac{m}{n}\right) = \frac{n!}{m!\,(n-m)!} \big(F(x)\big)^m \big(1 - F(x)\big)^{n-m}. \tag{10.3.4}$$

Die Verteilungsfunktion der in 10.2 definierten Zufallsvariablen $\zeta_k^{(n)}$ bezeichnen wir mit $\Phi_{kn}(x)$. Wir haben

$$\Phi_{kn}(x) = P(\zeta_k^{(n)} < x). \tag{10.3.5}$$

Wir wollen diese Verteilungsfunktion $\Phi_{kn}(x)$ durch $S_n(x)$ ausdrücken. Das Ereignis $\zeta_k^{(n)} < x$ besteht darin, daß in einer Gruppe von n Beobachtungen die (hinsichtlich der Größe) k-te Beobachtung kleiner als x ist, folglich also wenigstens k Beobachtungen kleiner als x sind. Dies ist gleichwertig damit, daß die beobachtete Häufigkeit $s_n(x)$ nicht kleiner als $\frac{k}{n}$ ist. So erhalten wir

$$\Phi_{kn}(x) = P(\zeta_k^{(n)} < x) = P\left(S_n(x) \geqq \frac{k}{n}\right) = \sum_{m=k}^{n} P\left(S_n(x) = \frac{m}{n}\right). \tag{10.3.6}$$

Aus den Gleichungen (10.3.4) und (10.3.6) ergibt sich

$$\Phi_{kn}(x) = \sum_{m=k}^{n} \frac{n!}{m!\,(n-m)!} \big(F(x)\big)^m \big(1 - F(x)\big)^{n-m}. \tag{10.3.7}$$

Diese Verteilungsfunktion kann man auch in der Form

$$\Phi_{kn}(x) = \frac{n!}{(k-1)!\,(n-k)!} \int_0^{F(x)} t^{k-1} (1-t)^{n-k}\, dt \tag{10.3.8}$$

schreiben.

Die Richtigkeit der Formel (10.3.8) ergibt sich durch Integration [man integriert mehrere Male partiell, wodurch man (10.3.7) erhält]. Die Einzelheiten dieses Beweises wollen wir übergehen.

Wir setzen nun voraus, daß die Dichte $f(x) = F'(x)$ der Zufallsvariablen X existiert. Dann existiert auch die Dichte $f_{kn}(x)$ der Zufallsvariablen $\zeta_k^{(n)}$. Wir differenzieren die Formel (10.3.8) nach x und erhalten

$$f_{kn}(x) = \frac{n!}{(k-1)!\,(n-k)!} \big(F(x)\big)^{k-1} \big(1 - F(x)\big)^{n-k} f(x). \tag{10.3.9}$$

D. Die Verteilungsfunktion $F(x)$ sei stetig. Wir leiten nun eine Beziehung zwischen den Verteilungsfunktionen $\Phi_{n-k+1,n}(x)$ und $\Phi_{kn}(x)$ oder zwischen den Verteilungen der Stichprobenfunktion $\zeta_{n-k+1}^{(n)}$ und der Stichprobenfunktion $\zeta_k^{(n)}$ her. Wir schreiben

$$\overline{F}(x) = P(-X < x) = P(X > -x) = 1 - F(-x)$$

und betrachten die unabhängigen Zufallsvariablen

$$-X_1, -X_2, \ldots, -X_n,$$

die alle dieselbe Verteilungsfunktion $\overline{F}(x) = 1 - F(-x)$ haben. Mit $\bar{\zeta}_k^{(n)}$ sei die Positionsfunktion der Zufallsvariablen $-X_1, -X_2, \ldots, -X_n$ bezeichnet. Sie wird auf dieselbe Weise wie die Positionsfunktion $\zeta_k^{(n)}$ der Zufallsvariablen $X_1, X_2, \ldots, X_n$ definiert. Ihre Verteilungsfunktion bezeichnen wir mit $\overline{\Phi}_{kn}(x)$.

Aus (10.3.8) erhalten wir

$$\overline{\Phi}_{kn}(x) = \frac{n!}{(k-1)!\,(n-k)!} \int\limits_0^{\overline{F}(x)} t^{k-1}(1-t)^{n-k}\,dt. \tag{10.3.10}$$

Daraus folgt

$$\overline{\Phi}_{kn}(-x) = \frac{n!}{(k-1)!\,(n-k)!} \int\limits_0^{1-F(x)} t^{k-1}(1-t)^{n-k}\,dt. \tag{10.3.11}$$

Aus (10.3.8) ergibt sich

$$\Phi_{n-k+1,n}(x) = \frac{n!}{(n-k)!\,(k-1)!} \int\limits_0^{F(x)} t^{n-k}(1-t)^{k-1}\,dt. \tag{10.3.12}$$

Im Integral (10.3.12) substituieren wir $y = 1 - t$ und erhalten

$$\begin{aligned}\Phi_{n-k+1,n}(x) &= -\frac{n!}{(n-k)!\,(k-1)!} \int\limits_1^{1-F(x)} (1-y)^{n-k} y^{k-1}\,dy \\ &= \frac{n!}{(n-k)!\,(k-1)!} \int\limits_{1-F(x)}^{1} y^{k-1}(1-y)^{n-k}\,dy.\end{aligned} \tag{10.3.13}$$

Wenn wir (10.3.11) und (10.3.13) miteinander vergleichen, erhalten wir

$$\Phi_{n-k+1,n}(x) = 1 - \overline{\Phi}_{kn}(-x). \tag{10.3.14}$$

Beispiel 10.3.1. Wir nehmen an, es sei

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

In einfachen Stichproben zu je zwei Elementen, die dieser normalen $N(0;1)$ Gesamtheit entnommen wurden, ordnen wir die beobachteten Werte der Größe nach an. In jeder Stichprobe kommen hier zwei Werte, x_1 und x_2, vor; es sei $x_1 < x_2$. Wir betrachten die Verteilung der Positionsfunktion $\zeta_1^{(2)}$.

Die Zufallsvariable $\zeta_1^{(2)}$ nimmt nun für jede Stichprobe den kleineren Wert x_1 an. Der Formel (10.3.9) entnehmen wir die Dichte der Zufallsvariablen $\zeta_1^{(2)}$:

$$f_{12}(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\frac{1}{\sqrt{2\pi}} \int\limits_x^{\infty} e^{-\frac{t^2}{2}}\,dt\right).$$

Die Dichte der Zufallsvariablen $\zeta_2^{(2)}$ ist durch folgende Formel gegeben:

$$f_{22}(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt\right).$$

Beispiel 10.3.2. Einer im Intervall [0, 1] gleichverteilten Gesamtheit, deren Dichte $f(x)$ durch

$$f(x) = \begin{cases} 1 & \text{für} \quad 0 \leqq x \leqq 1, \\ 0 & \text{für} \quad x < 0 \quad \text{und für} \quad x > 1 \end{cases}$$

gegeben ist, wollen wir einfache Stichproben zu je n Elementen entnehmen und die Verteilung der Positionsfunktion $\zeta_k^{(n)}$ bestimmen.

Ihre Dichte $f_{kn}(x)$ wird nach (10.3.9) durch

$$f_{kn}(x) = \begin{cases} \dfrac{n!}{(k-1)!\,(n-k)!} \left(\int\limits_0^x dt\right)^{k-1} \left(\int\limits_x^1 dt\right)^{n-k} \\ \quad = \dfrac{n!}{(k-1)!\,(n-k)!}\, x^{k-1}(1-x)^{n-k} & \text{für} \quad 0 < x < 1, \\ 0 & \text{für} \quad x \leqq 0 \quad \text{und für} \quad x \geqq 1. \end{cases}$$

dargestellt. Wenn wir diesen Ausdruck mit (5.9.3) vergleichen, so sehen wir, daß in diesem Beispiel die $\zeta_k^{(n)}$ eine Betaverteilung mit $p = k$ und $q = n - k + 1$ hat.

So sei zum Beispiel $n = 5$ und $k = 1$; dann hat $\zeta_1^{(5)}$ eine Betaverteilung mit $p = 1$ und $q = 5$. Wir berechnen die Wahrscheinlichkeit dafür, daß der Wert der Zufallsvariablen $\zeta_1^{(5)}$ zwischen 0,1 und 0,2 variiert. Es gilt

$$P(0{,}1 \leqq \zeta_1^{(5)} \leqq 0{,}2) = 5 \int\limits_{0,1}^{0,2} (1-x)^4\, dx = [-(1-x)^5]_{0,1}^{0,2} = -0{,}8^5 + 0{,}9^5 = 0{,}26281.$$

In unserem Beispiel war die unmittelbare Berechnung der gesuchten Wahrscheinlichkeit sehr einfach. Im allgemeinen kann man zur Berechnung von Wahrscheinlichkeiten einer Betaverteilung die Pearsonschen Tafeln (vgl. PEARSON [4]) benutzen.

10.4. Die stochastische Konvergenz einer Folge von Stichprobenquantilen

Von nun an wollen wir uns bei der Betrachtung von zentralen Positionsfunktionenfolgen $\{\zeta_k^{(n)}\}$ auf Folgen von Stichprobenquantilen beschränken, also auf den Fall $k = [n\lambda] + 1$. Bei der Untersuchung von äußeren Positionsfunktionenfolgen $\{\zeta_k^{(n)}\}$ wollen wir uns auf Folgen von Randelementen beschränken, also auf den Fall, daß k konstant oder aber gleich $n - l + 1$ ist, wobei l konstant ist.

Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien voneinander unabhängig und mögen dieselbe stetige Verteilungsfunktion $F(x)$ haben. Wir betrachten die Folge $\{\zeta_k^{(n)}\}$ von Stichprobenquantilen; also ist

$$k = [n\lambda] + 1,$$

wobei $0 < \lambda < 1$. Für jeden Wert λ aus dem Intervall $0 < \lambda < 1$ gibt es mindestens einen Wert a_λ derart, daß

$$1 - P(X = a_\lambda) \leqq F(a_\lambda) \leqq \lambda \tag{10.4.1}$$

ist. Wie man weiß, ist der Wert a_λ ein Lageparameter, nämlich das Verteilungsquantil jeder einzelnen der Zufallsvariablen X_k.

Jetzt wollen wir nachweisen, daß — falls (10.4.1) nur eine einzige Lösung a_λ besitzt — die Folge $\{\zeta_k^{(n)}\}$ der Stichprobenquantile stochastisch gegen das entsprechende Gesamtheitsquantil, also gegen a_λ, konvergiert. Speziell konvergiert daher die Folge der Stichprobenmedianen (siehe 10.2) stochastisch gegen die Mediane der Gesamtheit.

Satz 10.4.1. *Es bedeute $F(x)$ die Verteilungsfunktion der Zufallsvariablen X, ferner $\zeta_k^{(n)}$ das Quantil aus einfachen Stichproben* ($k = [n\lambda] + 1,\ 0 < \lambda < 1$). *Existiert nur ein Punkt $x = a_\lambda$, der der Ungleichung* (10.4.1) *genügt, so konvergiert die Folge der Positionsfunktionen $\zeta_k^{(n)}$ stochastisch gegen den Lageparameter a_λ.*

Beweis. Im Sinne der Definition der stochastischen Konvergenz (vgl. 6.2) müssen wir beweisen, daß für ein beliebiges $\varepsilon > 0$

$$\lim_{n\to\infty} P(|\zeta_k^{(n)} - a_\lambda| \geqq \varepsilon) = 0 \tag{10.4.2}$$

ist. Die Beziehung ist den zwei folgenden Beziehungen gleichwertig:

$$\lim_{n\to\infty} P(\zeta_k^{(n)} \geqq a_\lambda + \varepsilon) = 0,$$

$$\lim_{n\to\infty} P(\zeta_k^{(n)} \leqq a_\lambda - \varepsilon) = 0.$$

Aus der Definition 10.3.1 folgt, daß diese zwei Beziehungen den folgenden Beziehungen äquivalent sind:

$$\lim_{n\to\infty} P\left(S_n(a_\lambda + \varepsilon) < \frac{k}{n}\right) = 0, \tag{10.4.3}$$

$$\lim_{n\to\infty} P\left(S_n(a_\lambda - \varepsilon) \geqq \frac{k-1}{n}\right) = 0. \tag{10.4.3'}$$

Wir beachten jedoch, daß aus der Annahme, das der Beziehung (10.4.1) genügende a_λ sei eindeutig bestimmt, für hinreichend kleine $\varepsilon > 0$ die Beziehungen

$$\lim_{n\to\infty} \left(F(a_\lambda + \varepsilon) - \frac{k}{n}\right) = F(a_\lambda + \varepsilon) - \lambda = \delta_1 > 0, \tag{10.4.4}$$

$$\lim_{n\to\infty} \left(F(a_\lambda - \varepsilon) - \frac{k-1}{n}\right) = F(a_\lambda - \varepsilon) - \lambda = \delta_2 < 0 \tag{10.4.4'}$$

folgen. Wir zeigen nun, daß sich aus (10.4.4) die Beziehung (10.4.3) ergibt. In der Tat folgt aus (10.4.4) für hinreichend große n

$$F(a_\lambda + \varepsilon) - \frac{k}{n} > \frac{\delta_1}{2}.$$

Daraus folgt

$$P\left(S_n(a_\lambda + \varepsilon) < \frac{k}{n}\right) \leqq P\left(S_n(a_\lambda + \varepsilon) - F(a_\lambda + \varepsilon) < -\frac{\delta_1}{2}\right)$$

$$\leqq P\left(|S_n(a_\lambda + \varepsilon) - F(a_\lambda + \varepsilon)| > \frac{\delta_1}{2}\right). \qquad (10.4.5)$$

Da, wie wir aus 10.3 wissen, die Zufallsvariable $S_n(x)$ für feste x-Werte die Häufigkeit in einem Bernoullischen Versuchsschema mit der Erfolgswahrscheinlichkeit $p = F(x)$ ist, gilt im Punkt $x = a_\lambda + \varepsilon$

$$E(S_n(a_\lambda + \varepsilon)) = p = F(a_\lambda + \varepsilon), \qquad (10.4.6)$$

$$D^2(S_n(a_\lambda + \varepsilon)) = \frac{p(1-p)}{n} = \frac{F(a_\lambda + \varepsilon)(1 - F(a_\lambda + \varepsilon))}{n}. \qquad (10.4.7)$$

Aus der Tschebyscheffschen Ungleichung (3.3.4) und aus (10.4.5) erhalten wir

$$\lim_{n\to\infty} P\left(S_n(a_\lambda + \varepsilon) < \frac{k}{n}\right) \leqq \lim_{n\to\infty} \frac{4F(a_\lambda + \varepsilon)(1 - F(a_\lambda + \varepsilon))}{n\delta_1^2} \leqq \lim_{n\to\infty} \frac{1}{n\delta_1^2} = 0.$$

Damit haben wir die Gleichung (10.4.3) bewiesen.

Ganz analog kann man nachweisen, daß aus (10.4.4′) die Gleichung (10.4.3′) folgt. Der Satz 10.4.1 ist damit bewiesen.

10.5. Die Grenzverteilungen der Stichprobenquantile

Es sei $\zeta_k^{(n)}$ ein Stichprobenquantil, also $k = [n\lambda] + 1$ mit $0 < \lambda < 1$. Daher haben wir hier $\frac{k}{n} \to \lambda$ für $n \to \infty$. Ferner ist

$$\lim_{n\to\infty} \left(\frac{k}{n} - \lambda\right)\sqrt{n} = 0. \qquad (10.5.1)$$

Denn es ist $n\lambda < k = [n\lambda] + 1 \leqq \lambda n + 1$ und folglich

$$0 < \left(\frac{k}{n} - \lambda\right)\sqrt{n} \leqq \frac{1}{\sqrt{n}}.$$

Aus dieser Ungleichung erhalten wir (10.5.1).

Wir wollen nun annehmen, daß die Zufallsvariablen $X_1, X_2, \ldots, X_n$ unabhängig und stetig sind und alle dieselbe Verteilungsfunktion $F(x)$ haben. Die bloße Feststellung, daß die Folgen von Stichprobenquantilen stochastisch gegen die entsprechenden Lageparameter der Gesamtheit konvergieren, ist für die Anwendungen unzureichend. Sie erlaubt uns nämlich nicht, die Wahrscheinlichkeit dafür festzustellen, daß sich die Stichprobenquantile um eine bestimmte Größe von den entsprechenden Lageparametern unterscheiden.

Satz 10.5.1. *Es seien $F(x)$, $f(x)$ und $a_\lambda (0 < \lambda < 1)$ die Verteilungsfunktion, die Dichte bzw. das Quantil der Zufallsvariablen X, ferner sei $\zeta_k^{(n)}$ das Quantil aus einfachen Stichproben ($k = [n\lambda] + 1$), und $g_{kn}(y)$ sei die Dichte der Zufallsvariablen*

$$Y_k^{(n)} = \sqrt{\frac{n}{\lambda(1-\lambda)}}\, f(a_\lambda)\,(\zeta_k^{(n)} - a_\lambda). \tag{10.5.2}$$

Ist im Punkt $x = a_\lambda$ die Dichte $f(x)$ stetig und positiv, so gilt für jedes reelle Zahlenpaar y_1, y_2 mit $y_1 < y_2$ die Beziehung

$$\lim_{n\to\infty} P(y_1 < Y_k^{(n)} < y_2) = \lim_{n\to\infty} \int_{y_1}^{y_2} g_{kn}(y)\, dy = \frac{1}{\sqrt{2\pi}} \int_{y_1}^{y_2} e^{-\frac{y^2}{2}}\, dy. \tag{10.5.3}$$

Beweis. Die Dichte $f_{kn}(x)$ der Zufallsvariablen $\zeta_k^{(n)}$ hat nach (10.3.9) die Gestalt

$$f_{kn}(x) = \frac{n!}{(k-1)!\,(n-k)!} \big(F(x)\big)^{k-1} \big(1 - F(x)\big)^{n-k} f(x).$$

Nach (2.4.10) erhalten wir die Dichte $g_{kn}(y)$ in der Form

$$\begin{aligned} g_{kn}(y) &= \frac{1}{f(a_\lambda)} \sqrt{\frac{\lambda(1-\lambda)}{n}}\, f_{kn}\left(a_\lambda + \sqrt{\frac{\lambda(1-\lambda)}{n}} \frac{y}{f(a_\lambda)}\right) \\ &= \frac{1}{f(a_\lambda)} \sqrt{\frac{\lambda(1-\lambda)}{n}} \frac{n!}{(k-1)!\,(n-k)!} \big(F(x)\big)^{k-1} \big(1 - F(x)\big)^{n-k} f(x) \\ &= \sqrt{\frac{\lambda(1-\lambda)}{n}} \frac{n!}{(k-1)!\,(n-k+1)!} \lambda^{k-1} (1-\lambda)^{n-k+1} \frac{n-k+1}{1-\lambda} \\ &\quad \times \frac{f(x)}{f(a_\lambda)} \left(\frac{F(x)}{\lambda}\right)^{k-1} \left(\frac{1-F(x)}{1-\lambda}\right)^{n-k} \end{aligned} \tag{10.5.4}$$

mit

$$x = a_\lambda + \sqrt{\frac{\lambda(1-\lambda)}{n}} \frac{y}{f(a_\lambda)}.$$

Wir setzen

$$A_1(n) = \sqrt{\frac{\lambda(1-\lambda)}{n}} \frac{n!}{(k-1)!\,(n-k+1)!} \lambda^{k-1} (1-\lambda)^{n-k+1} \frac{n-k+1}{1\ \ \lambda},$$

$$A_2(n) = \frac{f\left(a_\lambda + \sqrt{\frac{\lambda(1-\lambda)}{n}} \frac{y}{f(a_\lambda)}\right)}{f(a_\lambda)}, \quad A_3(n) = \left(\frac{F(x)}{\lambda}\right)^{k-1} \left(\frac{1-F(x)}{1-\lambda}\right)^{n-k}.$$

Dann gilt die Gleichung

$$g_{kn}(y) = A_1(n)\, A_2(n)\, A_3(n). \tag{10.5.5}$$

Wir betrachten der Reihe nach die Faktoren $A_1(n)$, $A_2(n)$, $A_3(n)$.

Aus der Stirlingschen Formel $m! \approx \left(\frac{m}{e}\right)^m \sqrt{2\pi m}$ erhalten wir

$$\frac{n!}{(k-1)!\,(n-k+1)!} \approx \frac{n^n \sqrt{2\pi n}}{\sqrt{2\pi(k-1)}\,\sqrt{2\pi(n-k+1)}\,(k-1)^{k-1}(n-k+1)^{n-k+1}}.$$

Setzen wir näherungsweise $k-1 \approx n\lambda$ und $n-k+1 \approx n - n\lambda$, so folgt

$$\begin{aligned}\frac{n!}{(k-1)!\,(n-k+1)!} &\approx \frac{1}{\sqrt{2\pi\lambda(1-\lambda)\,n}}\,\frac{n^n}{(n\lambda)^{n\lambda}(n-n\lambda)^{n-n\lambda}}\\ &= \frac{1}{\sqrt{2\pi\lambda(1-\lambda)\,n}}\,\frac{1}{\lambda^{n\lambda}(1-\lambda)^{n-n\lambda}}\,\frac{n^n}{n^{n\lambda}\,n^{n-n\lambda}}\\ &= \frac{1}{\sqrt{2\pi\lambda(1-\lambda)\,n}}\,\frac{1}{\lambda^{k-1}(1-\lambda)^{n-k+1}}.\end{aligned}$$

Nach Substitution dieses Ausdrucks in die Formel für $A_1(n)$ erhalten wir

$$A_1(n) \approx \sqrt{\frac{\lambda(1-\lambda)}{n}}\,\frac{1}{\sqrt{2\pi\lambda(1-\lambda)\,n}}\,\frac{n-k+1}{1-\lambda} = \frac{1}{\sqrt{2\pi}}\,\frac{n-k+1}{n}\,\frac{1}{1-\lambda}.$$

Wegen $\frac{n-k+1}{n} \to 1-\lambda$ ergibt sich

$$\lim_{n\to\infty} A_1(n) = \frac{1}{\sqrt{2\pi}}. \tag{10.5.6}$$

Aus der Voraussetzung der Stetigkeit der Funktion $f(x)$ im Punkt $x = a_\lambda$ folgt die Gleichung

$$\lim_{n\to\infty} A_2(n) = 1. \tag{10.5.7}$$

Wir befassen uns nun mit dem Faktor $A_3(n)$. Die Funktion $F(x)$ können wir in der Umgebung des Punktes $x = a_\lambda$ folgendermaßen entwickeln:

$$\begin{aligned}F(x) &= F(a_\lambda) + F'(a_\lambda)\,(x-a_\lambda) + o(x-a_\lambda)\\ &= \lambda + \sqrt{\frac{\lambda(1-\lambda)}{n}}\,y + o\left(\sqrt{\frac{\lambda(1-\lambda)}{n}}\,\frac{y}{f(a_\lambda)}\right)\\ &= \lambda + y\sqrt{\frac{\lambda(1-\lambda)}{n}}\,\bigl(1 - Q(n,y)\bigr),\end{aligned} \tag{10.5.8}$$

wobei die Funktion $Q(n, y)$ für $n \to \infty$ gegen Null strebt.

Setzen wir den Ausdruck (10.5.8) in die Formel für $A_3(n)$ ein, dann erhalten wir

$$A_3(n) = \left\{1 + \frac{y\sqrt{\frac{\lambda(1-\lambda)}{n}}\left(1+Q(n,y)\right)}{\lambda}\right\}^{k-1} \left\{1 - \frac{y\sqrt{\frac{\lambda(1-\lambda)}{n}}\left(1+Q(n,y)\right)}{1-\lambda}\right\}^{n-k}.$$

Wir schreiben

$$a = \sqrt{\frac{1-\lambda}{\lambda}}\left(1+Q(n,y)\right), \qquad b = \sqrt{\frac{\lambda}{1-\lambda}}\left(1+Q(n,y)\right)$$

und erhalten

$$A_3(n) = \left(1 + \frac{ay}{\sqrt{n}}\right)^{k-1}\left(1 - \frac{by}{\sqrt{n}}\right)^{n-k}. \tag{10.5.9}$$

Daraus folgt

$$\log A_3(n) = (k-1)\log\left(1+\frac{ay}{\sqrt{n}}\right) + (n-k)\log\left(1-\frac{by}{\sqrt{n}}\right).$$

Da für $n \to \infty$ die Werte $\frac{a}{\sqrt{n}}$ und $\frac{b}{\sqrt{n}}$ gegen Null streben, kann man für hinreichend große n den letzten Ausdruck folgendermaßen entwickeln:

$$\begin{aligned}\log A_3(n) &= (k-1)\left(\frac{a}{\sqrt{n}}\,y - \frac{a^2}{2n}\,y^2 + o\left(\frac{1}{n}\right)\right)\\ &\quad + (n-k)\left(-\frac{b}{\sqrt{n}}\,y - \frac{b^2}{2n}\,y^2 + o\left(\frac{1}{n}\right)\right)\\ &= \frac{(k-1)\,a - (n-k)\,b}{\sqrt{n}}\,y - \frac{(k-1)\,a^2 + (n-k)\,b^2}{n}\,\frac{y^2}{2} + n\,o\left(\frac{1}{n}\right).\end{aligned} \tag{10.5.10}$$

Wir haben

$$\begin{aligned}\frac{(k-1)\,a - (n-k)\,b}{\sqrt{n}} &= \frac{(k-1)\sqrt{\frac{1-\lambda}{\lambda}} - (n-k)\sqrt{\frac{\lambda}{1-\lambda}}}{\sqrt{n}}\left(1+Q(n,y)\right)\\ &= \frac{(k-1)(1-\lambda) - (n-k)\lambda}{\sqrt{n}}\;\frac{1+Q(n,y)}{\sqrt{\lambda(1-\lambda)}}\\ &= \left(\left(\frac{k}{n}-\lambda\right)\sqrt{n} + \frac{\lambda-1}{\sqrt{n}}\right)\frac{1+Q(n,y)}{\sqrt{\lambda(1-\lambda)}}.\end{aligned}$$

Falls wir (10.5.1) berücksichtigen, erhalten wir

$$\lim_{n\to\infty} \frac{(k-1)\,a - (n-k)\,b}{\sqrt{n}} = 0. \tag{10.5.11}$$

Ferner haben wir

$$\begin{aligned}\lim_{n\to\infty} & \frac{(k-1)\,a^2 + (n-k)\,b^2}{n} \\ &= \lim_{n\to\infty} \left[\frac{k}{n}\,\frac{1-\lambda}{\lambda} + \left(1-\frac{k}{n}\right)\frac{\lambda}{1-\lambda} - \frac{a^2}{n}\right](1+Q(n,y))^2 \\ &= \lambda\,\frac{1-\lambda}{\lambda} + (1-\lambda)\,\frac{\lambda}{1-\lambda} = 1.\end{aligned} \tag{10.5.12}$$

Unter Beachtung der Formeln (10.5.10) bis (10.5.12) und der Tatsache, daß $n o\left(\frac{1}{n}\right) \to 0$ für $n \to \infty$ strebt, ergibt sich

$$\begin{aligned}&\lim_{n\to\infty} \log A_3(n) = -\frac{y^2}{2}; \\ &\lim_{n\to\infty} A_3(n) = e^{-\frac{y^2}{2}}.\end{aligned} \tag{10.5.13}$$

Aus den Formeln (10.5.6), (10.5.7) und (10.5.13) folgt sofort

$$\lim_{n\to\infty} g_{kn}(y) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}. \tag{10.5.14}$$

Da die Funktionen $A_1(n)$, $A_2(n)$, $A_3(n)$ in jedem Intervall (y_1, y_2) gleichmäßig beschränkt sind, erhalten wir schließlich aus (10.5.14)

$$\lim_{n\to\infty} P(y_1 < Y_k^{(n)} < y_2) = \lim_{n\to\infty} \int_{y_1}^{y_2} g_{kn}(y)\,dy = \int_{y_1}^{y_2} \lim_{n\to\infty} g_{kn}(y)\,dy = \frac{1}{\sqrt{2\pi}} \int_{y_1}^{y_2} e^{-\frac{y^2}{2}}\,dy.$$

Der Satz 10.5.1 ist damit bewiesen.

Aus Satz 10.5.1 folgt: Ist in der betrachteten Gesamtheit die Zufallsvariable X stetig mit der Dichte $f(x)$ verteilt und ist die Funktion $f(x)$ im Punkt $x = a_\lambda$ stetig und positiv, wobei a_λ ein Quantil der Gesamtheit ist, *dann ist das Quantil* $\zeta_k^{(n)}$ *aus einfachen Stichproben asymptotisch normal*

$$N\left(a_\lambda;\, \frac{1}{f(a_\lambda)} \sqrt{\frac{\lambda(1-\lambda)}{n}}\right) \tag{10.5.15}$$

verteilt.

Beispiel 10.5.1. Aus einer Grundgesamtheit, deren Merkmal X normal $N(0; 1)$ verteilt ist, erheben wir eine einfache Stichprobe vom Umfang $n = 121$. Die aus dieser Stichprobe errechnete Mediane sei $-0{,}5$. Wie groß ist die Wahrscheinlichkeit dafür, daß man aus einer Stichprobe vom gleichen Umfang eine kleinere Mediane als die beobachtete $-0{,}5$ erhält?

In diesem Beispiel ist $\lambda = \frac{1}{2}$, und die Mediane der Gesamtheit ist der Punkt $a_{\frac{1}{2}}$, der die Gleichung

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_{1/2}} e^{-\frac{x^2}{2}}\, dx = 0{,}5$$

erfüllt, also $a_{\frac{1}{2}} = 0$. Weiter ist $f\left(a_{\frac{1}{2}}\right) = \frac{1}{\sqrt{2\pi}}$.

Die Stichprobenmediane $\zeta_{61}^{(121)}$ ist asymptotisch normal

$$N\left(0; \frac{1}{2}\sqrt{\frac{2\pi}{121}}\right) = N(0; 0{,}114)$$

verteilt. Deshalb ist

$$P(\zeta_{61}^{(121)} \leqq -0{,}5) = P\left(\frac{\zeta_{61}^{(121)}}{0{,}114} \leqq \frac{-0{,}5}{0{,}114}\right) = \Phi(-4{,}39).$$

Aus den Tafeln der Normalverteilung lesen wir die gesuchte Wahrscheinlichkeit ab; sie ist äußerst klein, nämlich gleich 0,00001.

Beispiel 10.5.2. Wir betrachten eine Gesamtheit, in der das Merkmal X mit der Dichte

$$f(x) = \begin{cases} 2x & \text{für} \quad 0 \leqq x \leqq 1, \\ 0 & \text{für} \quad x < 0 \quad \text{und für} \quad x > 1 \end{cases}$$

verteilt ist.

Den Viertelwert $a_{\frac{1}{4}}$ dieser Gesamtheit bestimmen wir aus der Gleichung

$$\int_0^{a_{1/4}} 2x\, dx = 0{,}25;$$

es ist also $a_{\frac{1}{4}} = 0{,}5$ und $f\left(a_{\frac{1}{4}}\right) = 1$.

Aus dieser Gesamtheit erheben wir eine einfache Stichprobe vom Umfang $n = 70$ und bestimmen den Stichprobenviertelwert, d. h., wir beobachten den Wert der Positionsfunktion $\zeta_k^{(n)}$, wobei $n = 70$ und $k = [0{,}25 \cdot 70] + 1 = 18$ ist. Wie groß ist die Wahrscheinlichkeit dafür, einen Viertelwert nicht kleiner als 0,75 zu erhalten?

Die Positionsfunktion $\zeta_{18}^{(70)}$ hat in Annäherung die Normalverteilung

$$N\left(0{,}5; \sqrt{\frac{0{,}25 \cdot 0{,}75}{70}}\right) = N(0{,}5; 0{,}052).$$

Daraus erhalten wir

$$P(\zeta_{18}^{(70)} \geqq 0{,}75) = P\left(\frac{\zeta_{18}^{(70)} - 0{,}5}{0{,}052} \geqq \frac{0{,}25}{0{,}052}\right) \approx 1 - \Phi(4{,}81) \approx 0{,}0000.$$

Der Satz 10.5.1 ist ein Teil der allgemeinen Theorie der Grenzwertsätze für zentrale Folgen von Positionsstichprobenfunktionen, die von N. W. Smirnow [3] für den Fall entwickelt wurde, daß die $X_1, X_2, \ldots, X_n$ die gleiche Verteilung haben, während sie von Loève [5] und Cogburn [1] ohne Voraussetzung der gleichen Verteilung für die Variablen $X_1, X_2, \ldots, X_n$ hergeleitet wurde. Für eine Medianenfolge wurde der Satz 10.5.1 von Fogelson [1] bewiesen.

10.6. Die Grenzverteilungen der Randelemente einer Stichprobe

Am Ende von 10.2 definierten wir Zufallsvariable, die wir als Randelemente bezeichneten. Dies sind äußere Positionsfunktionen $\zeta_k^{(n)}$, für die k entweder konstant oder $k = n - l + 1$ mit konstantem l ist. Im ersten Fall ist $\lim\limits_{n\to\infty} \frac{k}{n} = 0$, während im zweiten Fall $\lim\limits_{n\to\infty} \frac{k}{n} = 1$ ist.

Wir wollen nun annehmen, daß die Zufallsvariablen $X_1, X_2, \ldots, X_n$ unabhängig und stetig mit derselben Dichte $f(x)$ verteilt sind.

Nach (10.3.9) ist die Dichte der Zufallsvariablen $\zeta_k^{(n)}$ durch

$$f_{kn}(x) = \frac{n!}{(k-1)!\,(n-k)!} \big(F(x)\big)^{k-1} \big(1 - F(x)\big)^{n-k} f(x)$$

gegeben.

Die Zufallsvariable $Z_k^{(n)}$ sei durch

$$Z_k^{(n)} = 2nF(\zeta_k^{(n)}) \tag{10.6.1}$$

definiert, sie nehme also nur Werte z aus dem Intervall $[0, 2n]$ an. Ihre Dichte bezeichnen wir mit $h_{kn}(z)$. Es gilt

$$h_{kn}(z) = \begin{cases} \dfrac{1}{2}\, \dfrac{(n-1)!}{(k-1)!\,(n-k)!} \left(\dfrac{z}{2n}\right)^{k-1} \left(1 - \dfrac{z}{2n}\right)^{n-k} & \text{für} \quad 0 \leqq z \leqq 2n, \\ 0 \qquad \text{für} \quad z < 0 \quad \text{und} \quad \text{für} \quad z > 2n. \end{cases} \tag{10.6.2}$$

Wir schreiben (10.6.2) in der Form

$$h_{kn}(z) = \frac{1}{(k-1)!}\, \frac{(n-1)!}{(n-k)!\, n^{k-1} \left(1 - \dfrac{z}{2n}\right)^k 2^k}\, z^{k-1} \left(1 - \frac{z}{2n}\right)^n.$$

Da k konstant ist, erhalten wir

$$\lim_{n\to\infty} \frac{(n-1)!}{(n-k)!\,n^{k-1}} = \lim_{n\to\infty} \frac{(n-k+1)\cdots(n-1)}{n^{k-1}} = \lim_{n\to\infty} \left(1-\frac{k-1}{n}\right)\cdots\left(1-\frac{1}{n}\right) = 1,$$

$$\lim_{n\to\infty} \frac{1}{\left(1-\frac{z}{2n}\right)^k} = 1.$$

Wenn wir beachten, daß $(k-1)! = \Gamma(k)$ und $\lim\limits_{n\to\infty}\left(1-\frac{z}{2n}\right)^n = e^{-\frac{z}{2}}$ ist, erhalten wir schließlich

$$\lim_{n\to\infty} h_{kn}(z) = \begin{cases} \dfrac{z^{k-1}e^{-\frac{z}{2}}}{2^k\,\Gamma(k)} & \text{für} \quad z \geqq 0. \\ 0 & \text{für} \quad z < 0. \end{cases} \tag{10.6.3}$$

Da die Funktionen $h_{kn}(z)$ für alle n in jedem endlichen Intervall $a < z < b$ gleichmäßig beschränkt sind, gilt

$$\lim_{n\to\infty} P(a < Z_k^{(n)} < b) = \lim_{n\to\infty} \int_a^b h_{kn}(z)\,dz$$

$$= \int_a^b \lim_{n\to\infty} h_{kn}(z)\,dz = \int_a^b \frac{z^{k-1}e^{-\frac{z}{2}}}{2^k\,\Gamma(k)}\,dz. \tag{10.6.4}$$

Die Zufallsvariable $Z_k^{(n)}$ hat somit eine asymptotische χ^2-Verteilung mit $m = 2k$ Freiheitsgraden. Ähnlich finden wir, daß die Zufallsvariable

$$U_k^{(n)} = 2n[1 - F(\zeta_k^{(n)})],$$

wobei $k = n - l + 1$ ist, eine asymptotische χ^2-Verteilung mit $2l$ Freiheitsgraden hat:

$$\lim_{n\to\infty} P(c < U_k^{(n)} < d) = \int_c^d \frac{u^{l-1}e^{-\frac{u}{2}}}{2^l\,\Gamma(l)}\,du. \tag{10.6.5}$$

Die Positionsfunktion $\zeta_k^{(n)}$ ist hier der l-te Stichprobenwert, vom größten Wert an gezählt.

Für die Bestimmung der asymptotischen Verteilung der Positionsfunktion $\zeta_k^{(n)}$ ist die Bedeutung der Formeln (10.6.4) und (10.6.5) nur gering, und zwar wegen der Schwierigkeiten, die bei der Bestimmung von $\zeta_k^{(n)}$ als Funktion von $Z_k^{(n)}$ oder $U_k^{(n)}$ auftreten. Als Anwendung dieser Formeln bringen wir nur ein einfaches Beispiel.

Beispiel 10.6.1. Die Zufallsvariable X habe die durch folgende Formel bestimmte Gleichverteilung:

$$f(x) = \begin{cases} 1 & \text{für} \quad 0 \leqq x \leqq 1, \\ 0 & \text{für} \quad x < 0 \quad \text{und für} \quad x > 1. \end{cases}$$

Wir bestimmen aus einfachen Stichproben die Positionsfunktion $\zeta_k^{(n)}$ bei festem k sowie die durch (10.6.1) definierte Zufallsvariable $Z_k^{(n)}$. In unserem Beispiel ist

$$Z_k^{(n)} = 2n\zeta_k^{(n)}.$$

Aus den vorherigen Betrachtungen folgt, daß die Zufallsvariable $2n\zeta_k^{(n)}$ eine asymptotische χ^2-Verteilung mit $2l$ Freiheitsgraden hat. Es sei zum Beispiel $n = 100$ und $k = 5$. Wir suchen die Wahrscheinlichkeit dafür, daß $\zeta_k^{(n)} > 0{,}0915$ ist. Es gilt

$$P(\zeta_k^{(n)} > 0{,}0915) = P(200\,\zeta_k^{(n)} > 18{,}3).$$

Aus den Tafeln der χ^2-Verteilung lesen wir ab, daß bei 10 Freiheitsgraden die gesuchte Wahrscheinlichkeit in Annäherung 0,05 beträgt.

Die Grenzverteilung des größten Stichprobenelements unter Voraussetzung der gleichen Verteilung für die Zufallsvariablen $X_1, X_2, \ldots, X_n$ untersuchten u. a. Fréchet [1], von Mises [2] und Gnedenko [8]. In [8] hat Gnedenko die Klasse aller Grenzverteilungen (bei entsprechender Normierung) des größten Stichprobenelements angegeben. N. W. Smirnow [3] wandte die Methode von Gnedenko bei der Untersuchung der Grenzverteilung eines beliebigen sukzessiven Stichprobenelements an. Mejzler [1] untersuchte die Grenzverteilungen des größten Stichprobenelements ohne die Voraussetzung, daß die Variablen $X_1, X_2, \ldots, X_n$ die gleiche Verteilung haben. Unter den gleichen Bedingungen entwickelten Loève [5] und Cogburn [1] eine allgemeine Theorie der Grenzverteilungen des sukzessiven Stichprobenelements sowie gemeinsame Grenzverteilungen von Gruppen sukzessiver Stichprobenelemente (siehe auch die Monographie von Gumbel [2]).

10.7. Die gemeinsame Verteilung einer Gruppe von Quantilen

Die stetigen Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien voneinander unabhängig und mögen alle dieselbe Verteilungsfunktion $F(x)$ und die Dichte $f(x)$ haben. Wir betrachten die Positionsstichprobenfunktionen

$$\zeta_1^{(n)}, \zeta_2^{(n)}, \ldots, \zeta_n^{(n)}.$$

Die Werte, die diese Positionsfunktionen annehmen können, bezeichnen wir entsprechend mit $y_1, y_2, \ldots, y_n$. Es soll die gemeinsame Verteilung der Zufallsvariablen

$$\left(\zeta_{j_1}^{(n)}, \zeta_{j_2}^{(n)}, \ldots, \zeta_{j_r}^{(n)}\right)$$

mit

$$1 \leqq j_1 < j_2 < \cdots < j_r \leqq n \quad \text{und} \quad 1 < r \leqq n$$

bestimmt werden. Wir teilen die x-Achse in die Intervalle

$$I_1 = (-\infty, y_{j_1}], \quad I_2 = (y_{j_1}, y_{j_2}], \quad \ldots, \quad I_{r+1} = (y_{j_r}, +\infty)$$

ein. Die gesuchte gemeinsame Wahrscheinlichkeitsverteilung finden heißt, für beliebige $(y_{j_1}, y_{j_2}, \ldots, y_{j_r})$ die Wahrscheinlichkeit des Ereignisses A zu bestimmen, das darin besteht, daß sich $j_1 - 1$ Werte von x im Intervall I_1 und ein Wert von x im „Intervall" $(y_{j_1}, y_{j_1} + dy_{j_1})$ befinden, weiter, daß sich $j_2 - j_1 - 1$ Werte von x im Intervall I_2 und einer im „Intervall"

$$(y_{j_2}, y_{j_2} + dy_{j_2})$$

befinden usw. und endlich, daß sich $n - j_r$ Werte von x im Intervall I_{r+1} befinden. Wir bezeichnen mit p_k die Wahrscheinlichkeit dafür, daß X einen Wert aus I_k annimmt:

$$p_k = \int_{I_k} f(x)\, dx \qquad (k = 1, 2, \ldots, r + 1).$$

Aus der Formel (2.3.8) folgt: Die Wahrscheinlichkeit dafür, daß X einen Wert aus dem „Intervall" $(y_{j_k}, y_{j_k} + dy_{j_k})$ annimmt, ist gleich $f(y_{j_k})\, dy_{j_k}$. Das Wahrscheinlichkeitselement der Zufallsvariablen

$$\left(\zeta_{j_1}^{(n)}, \zeta_{j_2}^{(n)}, \ldots, \zeta_{j_r}^{(n)}\right)$$

erhalten wir also aus (5.12.1′):

$$P(A) = n! \frac{p_1^{j_1-1} p_2^{j_2-j_1-1} \cdots p_r^{j_r-j_{r-1}-1} p_{r+1}^{n-j_r}}{(j_1-1)!\,1!\,(j_2-j_1-1)!\,1!\cdots(j_r-j_{r-1}-1)!\,1!\,(n-j_r)!} f(y_{j_1}) \cdots f(y_{j_r})\, dy_{j_1} \cdots dy_{j_r}$$

$$= g(y_{j_1}, \ldots, y_{j_r})\, dy_{j_1} \ldots dy_{j_r}. \tag{10.7.1}$$

Die Dichte der gemeinsamen Verteilung der Zufallsvariablen $\zeta_{j_1}^{(n)}, \ldots, \zeta_{j_r}^{(n)}$ ist also im Bereich $y_{j_1} < \cdots < y_{j_r}$ gleich $g(y_{j_1}, \ldots, y_{j_r})$, während sie für den übrigen Bereich gleich 0 ist.

Wir betrachten insbesondere die gemeinsame Verteilung der Zufallsvariablen $\left(\zeta_1^{(n)}, \zeta_n^{(n)}\right)$, d. h. die gemeinsame Verteilung des größten und des kleinsten x-Wertes in einer einfachen Stichprobe vom Umfang n. Für diesen Fall ist $r = 2$, $j_1 = 1$ und $j_2 = n$. Aus (10.7.1) erhalten wir für die gesuchte Dichtefunktion $g(y_1, y_n)$ die Formel

$$g(y_1, y_n) = \begin{cases} n(n-1)\left(\int\limits_{y_1}^{y_n} f(x)\, dx\right)^{n-2} f(y_1)\, f(y_n) & \text{für} \quad y_1 < y_n, \\ 0 & \text{für} \quad y_1 \geqq y_n. \end{cases} \tag{10.7.2}$$

10.8. Die Verteilung der Stichprobenbreite

In Kapitel 3 führten wir den Begriff der Variationsbreite ein, die ein Parameter der Verteilung der Grundgesamtheit ist und neben der Dispersion als Streuungsmaß dient.

Bei einer Stichprobe vom Umfang n, deren Elemente der Größe nach angeordnet sind, versteht man unter der *Stichprobenvariationsbreite* oder *Stichprobenbreite* die Differenz zwischen dem größten und dem kleinsten Stichprobenelement. Die Stichprobenbreite ist eine Zufallsvariable.

Definition 10.8.1. Die Zufallsvariable W, die durch

$$W = \zeta_n^{(n)} - \zeta_1^{(n)}$$

bestimmt ist, nennen wir *Stichprobenbreite*; dabei habe die Stichprobe den Umfang n.

Da nach Voraussetzung $\zeta_n^{(n)} \geqq \zeta_1^{(n)}$ ist, nimmt die Zufallsvariable W nur nichtnegative Werte an.

Die Verteilung der Stichprobenbreite (aus einfachen Stichproben) kann man aus (10.7.2) erhalten. Sie sind aber im allgemeinen kompliziert. Hier bemerken wir nur, daß für Stichproben aus einer normalen Gesamtheit die mit der Verteilung der Stichprobenbreite zusammenhängenden Wahrscheinlichkeiten in den Tafeln von Hartley und Pearson [1] und die Ausdrücke für die Momente in der Arbeit von Tippett [1] zusammengestellt sind.

Elfving [1] untersuchte die asymptotische Verteilung (für $n \to \infty$) der Breite einer einfachen Stichprobe aus einer normalen Gesamtheit; Gumbel [1] untersuchte die asymptotische Verteilung einer einfachen Stichprobe aus einer Gesamtheit, in der das betrachtete Merkmal X eine unbeschränkte Zufallsvariable ist, Momente beliebiger Ordnung besitzt und darüber hinaus eine beliebige Verteilung aufweisen kann.

Die Dichtefunktion $h(w)$ der Breite W einer n-gliedrigen einfachen Stichprobe, die aus einer Gesamtheit herausgegriffen wurde, in der X im Intervall $[0, 1]$ gleichverteilt ist, erhalten wir leicht aus den Formeln (6.5.7) und (10.7.2). In der Tat, (10.7.2) hat jetzt die Form

$$g(y_1, y_n) = \begin{cases} n(n-1)(y_n - y_1)^{n-2} & \text{für} \quad 0 \leqq y_1 < y_n \leqq 1, \\ 0 & \text{im übrigen Bereich.} \end{cases}$$

Setzen wir $w = y_n - y_1$ und erinnern wir uns daran, daß $0 \leqq y_1 + w \leqq 1$ und damit $0 \leqq y_1 \leqq 1 - w$ gilt, so erhalten wir

$$h(w) = \begin{cases} n(n-1) \int\limits_0^{1-w} w^{n-2}\, dy_1 = n(n-1)\, w^{n-2}(1-w) & \text{für} \quad 0 \leqq w \leqq 1, \\ 0 & \text{für} \quad w < 0,\ w > 1. \end{cases} \tag{10.8.2}$$

10.9. Die Toleranzgrenzen

Die Zufallsvariable X sei stetig mit der Dichte $f(x)$ verteilt. Es seien $\zeta_1^{(n)}, \ldots, \zeta_n^{(n)}$ Positionsstichprobenfunktionen, wobei die Stichproben vom Umfang n einfach sind und aus einer Gesamtheit stammen, deren Merkmal X wir untersuchen wollen. Wir betrachten das Integral

$$W = \int_{L_1}^{L_2} f(x)\,dx; \tag{10.9.1}$$

dabei sind $L_i = L_i(\zeta_1^{(n)}, \ldots, \zeta_n^{(n)})$ $(i = 1, 2)$ eindeutige Funktionen der Zufallsvariablen $\zeta_1^{(n)}, \ldots, \zeta_n^{(n)}$. Die Funktionen L_1 und L_2 sind Zufallsvariable, und daher ist auch das Integral W eine Zufallsvariable. Wir bemerken, daß W die Wahrscheinlichkeit für die Relation $L_1 \leqq X \leqq L_2$ ist. Diese Wahrscheinlichkeit ist wegen des zufälligen Charakters der Funktionen L_1 und L_2 eine Zufallsgröße. Mit anderen Worten: W ist der Anteil derjenigen Gesamtheitselemente, die zwischen den zufälligen Grenzen L_1 und L_2 liegen.

Definition 10.9.1. Die in (10.9.1) vorkommenden Funktionen L_1 und L_2 nennt man *Toleranzgrenzen*. Wenn die Verteilung der Zufallsvariablen W für alle Dichten $f(x)$ dieselbe ist, so nennen wir L_1 und L_2 *verteilungsunabhängige Toleranzgrenzen.*

Wir betrachten den Spezialfall $L_1 = \zeta_1^{(n)}$ und $L_2 = \zeta_n^{(n)}$ und zeigen, daß L_1 und L_2 verteilungsunabhängige Toleranzgrenzen sind. In der Tat, wir schreiben (10.9.1) in der Form

$$W = F(\zeta_n^{(n)}) - F(\zeta_1^{(n)}).$$

Aus der Voraussetzung, daß die Funktion $F(x)$ stetig ist, folgt (siehe 5.6), daß $Y = F(X)$ im Intervall $[0, 1]$ gleichverteilt ist. Abgesehen von den Bereichen, in denen die Funktion $F(x)$ konstant ist (dies hat keine Bedeutung), ist $F(x)$ eine wachsende Funktion, und damit ist W die Breite einer einfachen Stichprobe, die aus einer Grundgesamtheit mit Gleichverteilung herausgegriffen wurde. Die Dichtefunktion $h(w)$ der Zufallsvariablen W ist daher durch (10.8.2) gegeben. Diese Dichtefunktion ist von $F(x)$ unabhängig, und somit sind $\zeta_1^{(n)}$ und $\zeta_n^{(n)}$ von der Verteilung unabhängige Toleranzgrenzen. Man kann also für eine beliebige Zufallsvariable X mit stetiger Verteilungsfunktion und für ein beliebiges $\alpha > 0$ ein n so wählen, daß

$$P(W \geqq w) = n(n-1) \int_w^1 w^{n-2}(1-w)\,dw \geqq \alpha$$

gilt. Für ein geeignet gewähltes n haben wir daher eine Wahrscheinlichkeit $\geqq \alpha$ dafür, daß die Anzahl der Elemente der Gesamtheit, die zwischen dem kleinsten und dem größten Element einer n-gliedrigen Stichprobe enthalten sind,

zumindest den Wert w erreicht. Dieses Ergebnis wurde von WILKS [2] gewonnen. WALD [1] verallgemeinerte dieses Ergebnis auf mehrdimensionale Zufallsvariable.

Wir bemerken, daß W eine Betaverteilung (vgl. 5.9) mit $p = n - 1$ und $q = 2$ hat. Aus (5.9.5) und (5.9.6) erhalten wir somit

$$E(W) = \frac{n-1}{n+1}, \quad D^2(W) = \frac{n-1}{n+1} \, \frac{2}{(n+1)(n+2)}. \tag{10.9.2}$$

Hieraus erkennen wir, daß für relativ kleine n die Stichprobenfunktion W nahe bei 1 liegt.

Beispiel 10.9.1. Es sei $\alpha = 0{,}95$ und $w = 0{,}95$. Dann haben wir

$$P(W \geqq 0{,}95) = n(n-1) \int_{0{,}95}^{1} w^{n-2}(1-w)\,dw \geqq 0{,}95.$$

Hieraus erhalten wir $n \approx 93$.

Ein anderes für die Anwendungen wichtiges Beispiel für Toleranzengrenzen ist das folgende:

Beispiel 10.9.2. Es seien $\overline{X}$ und S^2 der Mittelwert bzw. die Varianz einer n-gliedrigen einfachen Stichprobe aus einer normalen Gesamtheit $N(m; \sigma)$, wobei m und σ unbekannt sind. Wir setzen

$$\begin{aligned} S' &= \sqrt{\frac{n}{n-1}}\, S, \\ W &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\overline{X}-hS'}^{\overline{X}+hS'} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx. \end{aligned} \tag{10.9.3}$$

Die Integrationsgrenzen $\overline{X} - hS'$ und $\overline{X} + hS'$ sind Zufallsvariable, und das gleiche gilt somit für W, die Anzahl der Elemente der Grundgesamtheit im Intervall $(\overline{X} - hS', \overline{X} + hS')$. Die Endpunkte dieses Intervalls sind die Toleranzgrenzen. Sie sind hier jedoch von der Verteilung nicht unabhängig. WALD und WOLFOWITZ [3] haben ein Verfahren zur angenäherten Berechnung der Wahrscheinlichkeit

$$P(W \geqq w) = \alpha$$

als einer Funktion von w, h und n angegeben, wobei W durch (10.9.3) gegeben ist. Tabellen, die mit der Berechnung dieser Wahrscheinlichkeiten zusammenhängen, wurden von BOWKER [1] bearbeitet.

Es muß abschließend noch betont werden, daß W für ein festes h eine Zufallsvariable ist; es ist daher nicht gerechtfertigt, W aus den Tafeln für die Normalverteilung zu berechnen, wie das häufig in der Praxis geschieht, wobei man $W = P(-h \leqq Y \leqq h)$ setzt und Y die Normalverteilung $N(0; 1)$ hat.

10.10. Der Satz von Gliwenko

A. In 10.3.A definierten wir für jeden Wert x die empirische Verteilungsfunktion $S_n(x)$. Entnehmen wir einer Grundgesamtheit eine Stichprobe vom Umfang n, so ist die Anzahl der beobachteten Stichprobenwerte, die kleiner als x sind, gleich $n S_n(x)$. Die Zufallsvariable $S_n(x)$ kann also nur Werte aus dem Intervall $[0, 1]$ annehmen. Wir stellten fest, daß die Funktion $S_n(x)$ nicht abnehmend und linksseitig stetig ist. Daraus folgt, daß $S_n(x)$ als Funktion von x dieselbenEigenschaften wie eine Verteilungsfunktion hat. Wir nannten sie deshalb eine *emprirische Verteilungsfunktion* (im Gegensatz zur Verteilungsfunktion $F(x)$ der Gesamtheit, die man auch die *theoretische* Verteilungsfunktion nennt). Wie wir schon erwähnten, ist in einfachen Stichproben für festes x die Zahl $S_n(x)$ die Häufigkeit in einem Bernoullischen Versuchsschema, also die Realisierungshäufigkeit eines zufälligen Ereignisses in n unabhängigen Versuchen, wobei die Realisierungswahrscheinlichkeit dieses Ereignisses in einem einzelnen Versuch gleich $p = F(x)$ ist.

Äußerst wichtig ist die Frage nach den Beziehungen, die zwischen den empirischen Verteilungsfunktionen $S_n(x)$ und der Verteilungsfunktion $F(x)$ der Gesamtheit bestehen. Es leuchtet ein, daß die Verteilungsfunktionen $S_n(x)$ und $F(x)$ sehr eng miteinander zusammenhängen.

Zunächst stellen wir fest, daß in einfachen Stichproben für jeden festen Wert x

$$P\left(\lim_{n\to\infty} S_n(x) = F(x)\right) = 1$$

gilt. Diese Beziehung folgt aus Satz 6.12.3, wenn wir beachten, daß für feste x-Werte der Ausdruck $S_n(x)$ die Häufigkeit in einem Bernoullischen Versuchsschema mit $p = F(x)$ ist.

Ein viel stärkeres Ergebnis enthält der Satz von GLIWENKO [1].

Satz 10.10.1 (GLIWENKO). *Es sei $S_n(x)$ eine empirische Verteilungsfunktion einer einfachen Stichprobe vom Umfang n, die einer Gesamtheit entnommen wurde, in der das Merkmal X die Verteilungsfunktion $F(x)$ hat. Die Wahrscheinlichkeit dafür, daß die Folge $\{S_n(x)\}$ für $n \to \infty$ gleichmäßig bezüglich x $(-\infty < x < \infty)$ gegen $F(x)$ konvergiert, ist gleich* 1.

Wir setzen

$$D_n = \sup_{-\infty<x<\infty} |S_n(x) - F(x)|. \tag{10.10.1}$$

Der Ausdruck D_n ist ebenfalls eine Zufallsvariable. *Der Satz von* GLIWENKO *besagt, daß die Relation*

$$P\left(\lim_{n\to\infty} D_n = 0\right) = 1 \tag{10.10.2}$$

besteht.

Dem Beweis des Gliwenkoschen Satzes lassen wir den Beweis des folgenden Lemmas vorangehen.

Lemma. *Es sei* $\{A_k\}$ $(k = 1, 2, \ldots)$ *eine Folge von zufälligen Ereignissen, und es sei*

$$P(A_k) = 1 \qquad (k = 1, 2, \ldots). \tag{10.10.3}$$

Dann gilt die Gleichung

$$P\Big(\prod_{k \geqq 1} A_k\Big) = 1.$$

Beweis des Lemmas. Zuerst beweisen wir die Gültigkeit des Lemmas für zwei Ereignisse A_1 und A_2. Ist nämlich $P(A_1) = P(A_2) = 1$, dann gilt

$$1 = P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2) = 1 + 1 - P(A_1 A_2).$$

Hieraus ergibt sich $P(A_1 A_2) = 1$. Mit Hilfe vollständiger Induktion übertragen wir dieses Ergebnis auf beliebig endlich viele Ereignisse; für $m = 2, 3, \ldots$ gilt

$$P(A_1 A_2 \cdots A_m) = 1. \tag{10.10.4}$$

Wir setzen jetzt

$$\prod_{k \geqq 1} A_k = A_1 (A_1 A_2) \cdot (A_1 A_2 A_3) \cdots = \prod_{m \geqq 1} B_m$$

mit $B_m = \prod_{k=1}^{m} A_k$. Wir sehen, daß für jedes m die Beziehung $B_{m+1} \subset B_m$ gilt; die Ereignisfolge $\{B_m\}$ ist also abnehmend. Aus Satz 1.3.4. erhalten wir

$$P\Big(\prod_{k \geqq 1} A_k\Big) = P\Big(\prod_{m \geqq 1} B_m\Big) = \lim_{m \to \infty} P(B_m). \tag{10.10.5}$$

Die Formeln (10.10.4) und (10.10.5) ergeben die Behauptung des Lemmas.

Beweis des Satzes von GLIWENKO. Es bedeute r eine natürliche Zahl. Mit $x_{r,k}$ $(k = 1, \ldots, r)$ sei der kleinste x-Wert, der der Relation

$$F(x) \leqq \frac{k}{r} \leqq F(x + 0) \tag{10.10.6}$$

genügt, bezeichnet. $S_n(x_{r,k})$ sei die empirische Verteilungsfunktion einer einfachen Stichprobe vom Umfang n an der Stelle $x_{r,k}$, und A_k^r bzw. A_0^r $(k = 1, 2, \ldots, r - 1)$ sei das Ereignis, das darin besteht, daß die Relation

$$\lim_{n \to \infty} \max\,[|S_n(x_{r,1}) - F(x_{r,1})|,\ |S_n(x_{r,r} + 0) - F(x_{r,r} + 0)|] = 0$$

bzw. die Relation

$$\lim_{n \to \infty} \max\,[|S_n(x_{r,k} + 0) - F(x_{r,k} + 0)|,\ |S_n(x_{r,k+1}) - F(x_{r,k+1})|] = 0$$

gilt. Aus Satz 6.12.3 folgt

$$P(A_k^r) = 1 \qquad (k = 0, 1, 2, \ldots, r-1). \tag{10.10.7}$$

A^r bedeute den Durchschnitt $A_0^r A_1^r \cdots A_{r-1}^r$. Das Ereignis A^r ist dem Ereignis

$$\lim_{n\to\infty} \max_{1\leqq k\leqq r} \max\,[|S_n(x_{r,k}) - F(x_{r,k})|, \quad |S_n(x_{r,k}+0) - F(x_{r,k}+0)|] = 0$$

gleichwertig.

Aus (10.10.7) und der Definition des Ereignisses A^r folgt nach dem Lemma die Gleichung

$$P(A^r) = 1.$$

Wir setzen $A = A^1 A^2 A^3 \cdots$. Im Sinne des Lemmas erhalten wir

$$P(A) = P\Big(\prod_{r\geqq 1} \lim_{n\to\infty} \max_{1\leqq k\leqq r} \max\,[|S_n(x_{r,k}) - F(x_{r,k})|, \\ |S_n(x_{r,k}+0) - F(x_{r,k}+0)|] = 0\Big) = 1. \tag{10.10.8}$$

Nun sei x eine beliebige Zahl, die der Ungleichung $x_{r,k} < x \leqq x_{r,k+1}$ genügt. Es gilt

$$S_n(x_{r,k}+0) \leqq S_n(x) \leqq S_n(x_{r,k+1}), \qquad F(x_{r,k}+0) \leqq F(x) \leqq F(x_{r,k+1}).$$

Hieraus erhalten wir

$$S_n(x_{r,k}+0) - F(x_{r,k+1}) \leqq S_n(x) - F(x) \leqq S_n(x_{r,k+1}) - F(x_{r,k}+0). \tag{10.10.9}$$

Da für $x_{r,k} < x_{r,k+1}$ aus (10.10.6)

$$0 \leqq F(x_{r,k+1}) - F(x_{r\,k}+0) \leqq \frac{1}{r}$$

folgt, ergibt sich aus (10.10.9) für $k = 1, \ldots, r-1$ die Beziehung

$$\sup_{x_{r,k}<x\leqq x_{r,k+1}} |S_n(x) - F(x)| \leqq \max\,[|S_n(x_{r,k+1}) - F(x_{r,k+1})|, \\ |S_n(x_{r,k}+0) - F(x_{r,k}+0)|] + \frac{1}{r}.$$

Daraus folgt die Relation

$$\max\Big[\sup_{x_{r,1}\leqq x\leqq x_{r,r}} |S_n(x) - F(x)|, \quad |S_n(x_{r,r}+0) - F(x_{r,r}+0)|\Big] \\ \leqq \max_{1\leqq k\leqq r} \max\,[|S_n(x_{r,k}) - F(x_{r,k})|, \quad |S_n(x_{r,k}+0) - F(x_{r,k}+0)|] + \frac{1}{r}. \tag{10.10.10}$$

Aus (10.10.10) erhalten wir

$$P\left(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |S_n(x) - F(x)| = 0\right)$$

$$\geqq P\left(\lim_{n\to\infty} \lim_{r\to\infty} \max_{1\leqq k\leqq r} \max\left[|S_n(x_{r,k}) - F(x_{r,k})|,\ |S_n(x_{r,k}+0) - F(x_{r,k}+0)|\right] = 0\right)$$

$$\geqq P\left(\prod_{r\geqq 1} \lim_{r\to\infty} \max_{1\leqq k\leqq r} \max\left[|S_n(x_{r,k}) - F(x_{r,k})|,\ |S_n(x_{r,k}+0) - F(x_{r,k}+0)|\right] = 0\right).$$

Dies und die Relation (10.10.8) führt schließlich zu

$$P\left(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |S_n(x) - F(x)| = 0\right) = 1.$$

Der Satz von GLIWENKO ist damit bewiesen.

B. Die empirische Verteilungsfunktion kann man auch in der folgenden Form schreiben: Es sei $\{X_r\}$ $(r = 1, 2, \ldots)$ eine Folge von Zufallsvariablen. Wir setzen für jedes x $(-\infty < x < \infty)$

$$f_x(X_r) = \begin{cases} 0 & \text{für} \quad X_r \geqq x, \\ 1 & \text{für} \quad X_r < x. \end{cases} \tag{10.10.11}$$

Dann haben wir

$$S_n(x) = \frac{1}{n} \sum_{r=1}^{n} f_x(X_r). \tag{10.10.12}$$

Der Satz von GLIWENKO wurde in verschiedenen Richtungen verallgemeinert. WOLFOWITZ [5], [6], [8] verallgemeinerte ihn für den Fall, daß nicht alle Zufallsvariablen X_r die gleiche Verteilungsfunktion haben, sowie auf Zufallsvektoren. Den zuletzt genannten Fall betrachtete auch BLUM [1]. FORTET und MOURIER [1] sowie FISZ [5] verallgemeinerten den Satz von GLIWENKO auf den Fall, daß in Formel (10.10.12) die Funktionen $f_x(X_r)$ allgemeiner sind als die durch (10.10.11) definierten Funktionen. TUCKER [1] verallgemeinerte ihn auf den Fall, daß die Folge der Zufallsvariablen X_r stationär im engeren Sinne ist (siehe auch VARADARAJAN [3] sowie RANGA-RAO [1]).

10.11. Die Sätze von Kolmogoroff und Smirnow

A. Die Grenzverteilung der Stichprobenfunktion D_n, die durch die Formel (10.10.1) definiert wurde, bildet den Gegenstand des nun folgenden Satzes. Wir bezeichnen die Verteilungsfunktion der Zufallsvariablen $\sqrt{n}D_n$ mit

$$Q_n(\lambda) = \begin{cases} P\left(D_n\sqrt{n} < \lambda\right) = P\left(D_n < \dfrac{\lambda}{\sqrt{n}}\right) & \text{für} \quad \lambda > 0, \\ 0 & \text{für} \quad \lambda \leqq 0. \end{cases} \tag{10.11.1}$$

Satz 10.11.1. *Es sei $S_n(x)$ eine empirische Verteilungsfunktion einer einfachen Stichprobe vom Umfang n, die einer Gesamtheit entnommen wurde, in der das Merkmal X eine stetige Verteilungsfunktion $F(x)$ besitzt. Dann gilt die Beziehung*

$$\lim_{n\to\infty} Q_n(\lambda) = Q(\lambda) = \begin{cases} \sum\limits_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2\lambda^2) & \text{für } \lambda > 0, \\ 0 & \text{für } \lambda \leqq 0. \end{cases} \tag{10.11.2}$$

Dies ist der Satz von KOLMOGOROFF [6]; eine Idee des Beweises geben wir in 10.11.C an.

In Tafel VIII am Ende des Buches sind die Werte von $Q(\lambda)$ für einige Werte von λ zusammengestellt. Diese Daten entnahmen wir den Tafeln für $Q(\lambda)$ von SMIRNOW [1].

Tabellen der Verteilungsfunktion $Q_n(\lambda)$ für kleine n kann man in der Arbeit von BIRNBAUM [1] finden.

Wir bemerken, daß die Grenzverteilung $Q(\lambda)$ von der theoretischen Verteilungsfunktion $F(x)$ unabhängig ist, von der wir lediglich voraussetzen, daß sie stetig ist. Von diesem Gesichtspunkt aus sagen wir, daß D_n eine *verteilungsfreie* Grenzverteilung besitzt. (Eine allgemeine Charakteristik der verteilungsfreien Stichprobenfunktionen findet man in den Arbeiten von BIRNBAUM [2], BIRNBAUM und RUBIN [1] sowie BELL [1].) Ist jedoch die theoretische Verteilungsfunktion $F(x)$ nicht stetig, so gilt die Beziehung (10.11.2) nicht. Diesen Fall hat SCHMID [1] untersucht, der bewiesen hat, daß in diesem Fall der Grenzwert der Funktion $Q_n(\lambda)$ von der theoretischen Verteilungsfunktion abhängig ist, und zwar von den Werten von $F(x)$ in den Unstetigkeitspunkten (siehe auch GICHMAN [3]).

Beispiel 10.11.1. Es liege ein Posten Schrauben vor, deren Durchmesser gleich sein sollen. Genaue Messungen der Durchmesser zeigen jedoch einige Unstimmigkeiten; dabei wollen wir aber annehmen, daß bei der Produktion dieser Schrauben keine wesentlichen Abweichungen von der festgesetzten Norm vorkommen. Die Meßergebnisse x, die wir als Werte der Zufallsvariablen X betrachten, sind dann im allgemeinen normal verteilt. So sei zum Beispiel

$$f(x) = \frac{3}{\sqrt{2\pi}} \exp\left(-\frac{9}{2}(x-2)^2\right) \tag{10.11.3}$$

die Dichte dieser Normalverteilung, d. h., die Zufallsvariable X hat die Verteilung $N(2; 1/3)$, wobei ein Millimeter als Einheit genommen ist.

Diesem Schraubenposten entnehmen wir eine einfache Stichprobe vom Umfang $n = 80$. Wir stellen also 80 Messungen an. Die Meßergebnisse seien der Größe nach angeordnet, also $x_1 < x_2 < \cdots < x_{80}$. Wir berechnen die obere Schranke der absoluten Differenz $|s_{80}(x) - F(x)|$, wobei $F(x)$ die Verteilungsfunktion der Zufallsvariablen X ist, während $s_{80}(x)$ die empirische Verteilungsfunktion der Stichprobe darstellt. In unserer Stichprobe haben wir den Wert $d_{80} = 0{,}21$ beobachtet. Da n ziemlich groß ist, benutzen wir den Kolmogoroffschen Satz und berechnen die Wahrscheinlichkeit dafür, daß $D_{80}(x)$ nicht kleiner als der beobachtete Wert ist.

Wir haben $d_{80}\sqrt{80} = 0{,}21 \cdot 8{,}94 = 1{,}88 = \lambda$. In Tafel VIII finden wir

$$P(D_{80} \geqq 0{,}21) = P\left(D_{80} \geqq \frac{1{,}88}{8{,}94}\right) \approx 1 - Q(1{,}88) \approx 0{,}0017.$$

Diese Wahrscheinlichkeit ist von der Ordnung 0,001 und könnte in uns den Argwohn wecken, daß die Zufallsvariable X gar nicht nach der Formel (10.11.3) verteilt ist oder daß die Stichprobenauswahl tatsächlich nicht nur vom Zufall beherrscht war. Die Methode, derartige Schlüsse zu ziehen, besprechen wir in den nächsten Kapiteln.

B. Wir gehen jetzt zum Satz von Smirnow über.

Die Zufallsvariablen $X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, X_{22}, \ldots, X_{2n_2}$ seien voneinander unabhängig und mögen dieselbe Verteilungsfunktion $F(x)$ haben. $S_{1n_1}(x)$ und $S_{2n_2}(x)$ seien die empirischen Verteilungsfunktionen der Stichproben $(x_{11}, \ldots, x_{1n_1})$ bzw. $(x_{21}, \ldots, x_{2n_2})$. Damit sind $S_{1n_1}(x)$ und $S_{2n_2}(x)$ empirische Verteilungsfunktionen für zwei unabhängige einfache Stichproben aus derselben Gesamtheit, in der das Merkmal X eine Zufallsvariable mit der Verteilungsfunktion $F(x)$ ist; n_1 bzw. n_2 ist der Umfang der ersten bzw. der zweiten Stichprobe. Wir betrachten die Zufallsvariablen

$$D^+_{n_1,n_2} = \max_{-\infty<x<\infty} \left(S_{1n_1}(x) - S_{2n_2}(x)\right), \tag{10.11.4}$$

$$D_{n_1,n_2} = \max_{-\infty<x<\infty} |S_{1n_1}(x) - S_{2n_2}(x)|. \tag{10.11.5}$$

Mit der Bezeichnung $n = \dfrac{n_1 n_2}{n_1 + n_2}$ seien $Q^+_{n_1,n_2}(\lambda)$ und $Q_{n_1,n_2}(\lambda)$ die Verteilungsfunktionen der Zufallsvariablen $\sqrt{n}\,D^+_{n_1,n_2}$ bzw. $\sqrt{n}\,D_{n_1,n_2}$:

$$Q^+_{n_1,n_2}(\lambda) = \begin{cases} P\left(D^+_{n_1,n_2} < \dfrac{\lambda}{\sqrt{n}}\right) & \text{für } \lambda > 0, \\ 0 & \text{für } \lambda \leqq 0, \end{cases} \tag{10.11.6}$$

$$Q_{n_1,n_2}(\lambda) = \begin{cases} P\left(D_{n_1,n_2} < \dfrac{\lambda}{\sqrt{n}}\right) & \text{für } \lambda > 0, \\ 0 & \text{für } \lambda \leqq 0. \end{cases} \tag{10.11.7}$$

Es gilt nun der folgende Satz von Smirnow.

Satz 10.11.2. *$S_{1n_1}(x)$ und $S_{2n_2}(x)$ seien die empirischen Verteilungsfunktionen zweier unabhängiger einfacher Stichproben, die den Umfang n_1 bzw. n_2 haben. Diese Stichproben seien einer Gesamtheit entnommen, in der das Merkmal X eine*

stetige Verteilungsfunktion hat. Es gelten dann die Relationen

$$\lim_{\substack{n_1\to\infty\\ n_2\to\infty}} Q^+_{n_1,n_2}(\lambda) = \begin{cases} 1 - \exp(-2\lambda^2) & \textit{für } \lambda > 0, \\ 0 & \textit{für } \lambda \leqq 0; \end{cases} \tag{10.11.8}$$

$$\lim_{\substack{n_1\to\infty\\ n_2\to\infty}} Q_{n_1,n_2}(\lambda) = \begin{cases} \sum\limits_{j=-\infty}^{\infty} (-1)^j \exp(-2j^2\lambda^2) & \textit{für } \lambda > 0, \\ 0 & \textit{für } \lambda \leqq 0. \end{cases} \tag{10.11.9}$$

Wir bemerken noch, daß SMIRNOW diesen Satz unter der zusätzlichen Voraussetzung

$$\frac{n_1}{n_2} = \tau \qquad (\tau > 0)$$

bewies. GICHMAN [1] stellte jedoch fest, daß diese Voraussetzung überflüssig ist.

C. Interessant ist die Geschichte der Beweise des Kolmogoroffschen und des Smirnowschen Satzes. Die ursprünglichen Beweise waren sehr kompliziert. Einen einfacheren, jedoch noch ziemlich schwerfälligen Beweis erbrachte FELLER [5]. Einen kurzen und eleganten Beweis, der sich auf eine heuristische Annahme stützte, gab DOOB [4] an, die Richtigkeit dieser heuristischen Annahme wies DONSKER [2] nach.

Wir beschränken uns hier darauf, die Idee des Doobschen Beweises von Satz 10.11.1 anzugeben, und setzen dazu

$$U_n(x) = \sqrt{n}\big(S_n(x) - F(x)\big).$$

Da die Verteilungsfunktion $F(x)$ stetig ist, können wir annehmen, daß die Zufallsvariable X im Intervall $[0, 1]$ gleichmäßig verteilt ist; anderenfalls betrachten wir die Zufallsvariable $Y = F(X)$, die eine ebensolche gleichmäßige Verteilung (vgl. 5.6) besitzt. Wir können also

$$U_n(x) = \sqrt{n}\big(S_n(x) - x\big) \qquad (0 \leqq x \leqq 1),$$

$$D_n\sqrt{n} = \sup_{0\leqq x\leqq 1} |U_n(x)|$$

schreiben und leicht die Beziehungen

$$\left.\begin{aligned} &E\big(U_n(x)\big) = 0 && (0 \leqq x \leqq 1), \\ &E\big(U_n(x_1)U_n(x_2)\big) = x_1(1 - x_2) && (0 \leqq x_1 \leqq x_2 \leqq 1) \end{aligned}\right\} \tag{*}$$

finden. Aus dem mehrdimensionalen Moivre-Laplaceschen Satz 6.13.1 folgt, daß für $r = 1, 2, 3, \ldots$ und beliebige Punkte $x_1, \ldots, x_r$ $(0 \leqq x_1 \leqq \cdots \leqq x_r \leqq 1)$

der Vektor $\big(U_n(x_1), \ldots, U_n(x_r)\big)$ asymptotisch normal verteilt ist, wobei die Mittelwerte, Dispersionen und Kovarianzen durch die Formeln (*) gegeben sind.

Wir betrachten jetzt einen normalen stochastischen Prozeß $\{U(x), 0 \leqq x \leqq 1\}$, d. h. einen solchen Prozeß, für den für $r = 1, 2, 3, \ldots$ und für beliebige $x_1, \ldots, x_r$ $(0 \leqq x_1 \leqq \cdots \leqq x_r \leqq 1)$ der Vektor $(U(x_1), \ldots, U(x_r))$ eine Normalverteilung besitzt, deren Mittelwerte, Dispersionen und Kovarianzen durch die Formeln (*) gegeben sind. (Die Realisierungen dieses Prozesses sind mit der Wahrscheinlichkeit 1 stetig.)

Die heuristische Voraussetzung von Doob besagt, daß die Relation

$$\lim_{n\to\infty} P\left(\sup_{0\leqq x\leqq 1} |U_n(x)| < \lambda\right) = P\left(\max_{0\leqq x\leqq 1} |U_n(x)| < \lambda\right)$$

besteht. Doob [4] stellte fest, daß die rechte Seite dieser Gleichung gleich der Funktion $Q(\lambda)$ ist, die durch die Formel (10.11.2) gegeben ist.

Auch den Satz von Smirnow kann man mit Hilfe der Doobschen Methode beweisen. Wir wollen jedoch einen anderen vollständigen Beweis dieses Satzes für den Spezialfall $n_1 = n_2 = m$, also $n = \frac{m}{2}$ bringen.

Der Beweis verläuft folgendermaßen: Wir leiten zuerst ein Lemma her, das die exakten Verteilungen der Stichprobenfunktionen $\sqrt{\frac{m}{2}}\, D^+_{m,m}$ und $\sqrt{\frac{m}{2}}\, D_{m,m}$ angibt, gehen dann mit $m \to \infty$ zur Grenze über und erhalten so den Satz 10.11.2. Diesen einfachen und scharfsinnigen Beweis erbrachten Gnedenko und Koroljuk in der Arbeit [1], die eine Reihe von wichtigen Untersuchungen auf diesem Gebiet auslöste, deren Besprechung der Leser in der Arbeit von Gnedenko [7] findet. Später zeigte Koroljuk [1], daß die von ihm und Gnedenko angegebene Beweismethode sich auch auf den Beweis des Smirnowschen Satzes in der allgemeinen Form übertragen läßt; sie findet auch beim Beweis des Kolmogoroffschen Satzes Anwendung.

Lemma. *Die Voraussetzungen des Satzes* 10.11.2 *seien erfüllt. Es gelten dann die Relationen*

$$Q^+_{m,m}(\lambda) = \begin{cases} 0 & \textit{für} \quad \lambda \leqq 0, \\ 1 - \dfrac{\binom{2m}{m-c}}{\binom{2m}{m}} & \textit{für} \quad 0 < \lambda \leqq \sqrt{\dfrac{m}{2}}, \\ 1 & \textit{für} \quad \lambda > \sqrt{\dfrac{m}{2}}, \end{cases} \tag{10.11.10}$$

$$Q_{m,m}(\lambda) = \begin{cases} 0 & \text{für } \lambda \leqq \frac{1}{\sqrt{2m}}, \\ \sum_{j=-[m/c]}^{[m/c]} (-1)^j \frac{\binom{2m}{m-jc}}{\binom{2m}{m}} & \text{für } \frac{1}{\sqrt{2m}} < \lambda \leqq \sqrt{\frac{m}{2}}, \\ 1 & \text{für } \lambda > \sqrt{\frac{m}{2}}, \end{cases} \qquad (10.11.11)$$

wobei $c = -\left[-\lambda\sqrt{2m}\right]$ *ist.*

Beweis des Lemmas. Mit $y_1, y_2, \ldots, y_{2m}$ bezeichnen wir die der Größe nach angeordneten Werte $x_{11}, x_{12}, \ldots, x_{1m}, x_{21}, x_{22}, \ldots, x_{2m}$, so daß $y_1 < y_2 < \cdots < y_{2m}$ gilt. Wir führen die Zufallsvariablen

$$Z_j \ (j = 1, 2, \ldots, 2m)$$

ein, die die Werte 1 oder -1 annehmen, je nachdem, ob y_i ein Element der ersten oder der zweiten Stichprobe ist. Ferner setzen wir

$$V_0 = 0, \qquad V_k = \sum_{j=1}^{k} Z_j \qquad (k = 1, 2, \ldots, 2m).$$

Es gelten hier die Gleichungen

$$m D^+_{m,m} = m \max_{-\infty<x<\infty} (S_{1m}(x) - S_{2m}(x)) = \max_{0\leqq k\leqq 2m} V_k, \qquad (10.11.12)$$

$$m D_{m,m} = m \max_{-\infty<x<\infty} |S_{1m}(x) - S_{2m}(x)| = \max_{0\leqq k\leqq 2m} |V_k|. \qquad (10.11.13)$$

Die Gleichung (10.11.12) begründet man so: Es ist $m\big(S_{1m}(x) - S_{2m}(x)\big)$ für ein festes x die Differenz zwischen der Anzahl der Elemente der ersten Stichprobe, die kleiner als x sind, und der Anzahl ebensolcher Elemente der zweiten Stichprobe, $m D^+_{m,m}$ bezeichnet die größte dieser Differenzen, die wir erhalten, wenn x das Intervall $(-\infty, \infty)$ durchläuft. Beachten wir die Definition der Größen Z_j und V_k, so erhalten wir die Gleichung (10.11.12). Ähnlich wird (10.11.13) hergeleitet.

Den weiteren Gedankengang erleichtert die folgende Illustration: Wir betrachten die Irrfahrt eines Teilchens auf der t, v-Ebene. Im Anfangsmoment $t = 0$ befinde sich das Teilchen im Zustand $v = 0$. Das Teilchen unterliegt zufälligen Einflüssen in den Augenblicken $t = 1, 2, \ldots, 2m$, in deren Folge sich im Augenblick $j\,(1 \leqq j \leqq 2m)$ seine Ordinate v um 1 oder -1 ändert, je nachdem, ob $Z_j = 1$ oder $Z_j = -1$ ist. Dabei befinden sich unter den $2m$ Verschiebungen

m positive und m negative, so daß im Augenblick $t = 2m$ wieder $v = 0$ ist. Wir verbinden die Punkte mit den Koordinaten (t, v) durch einen Streckenzug, den wir *Trajektorie* nennen. Dieser Streckenzug ist vom Zufall abhängig, sein Anfang ist

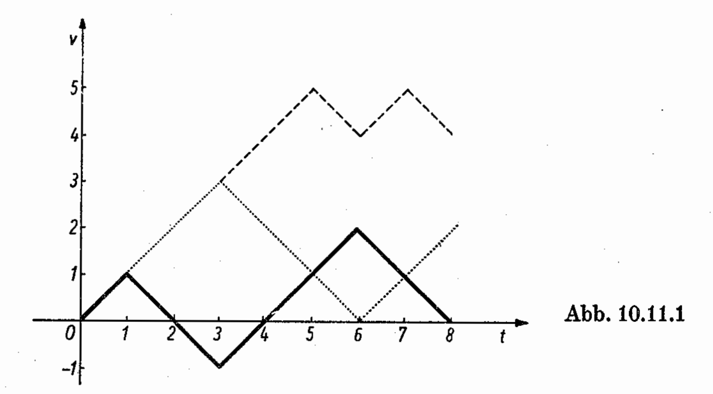


Abb. 10.11.1

der Punkt (0,0) und sein Ende der Punkt $(2m, 0)$. Eine derartige Trajektorie wird durch die dick ausgezeichnete Linie in Abb. 10.11.1 dargestellt.

Aus Formel (10.11.6) erhalten wir

$$Q^+_{m,m}(\lambda) = P\left(\sqrt{\frac{m}{2}}\, D^+_{m,m} < \lambda\right) = P\left(m D^+_{m,m} < \lambda \sqrt{2m}\right).$$

Da $m D^+_{m,m}$ nur ganze Zahlen annimmt, erhalten wir unter Berücksichtigung von (10.11.12)

$$Q^+_{m,m}(\lambda) = P\left(\max_{1 \leq k \leq 2m} V_k < c\right); \tag{10.11.14}$$

dabei kann c eine beliebige natürliche Zahl sein.

Die Berechnung der Verteilungsfunktion $Q^+_{m,m}(\lambda)$ wird hier also auf die Berechnung der Wahrscheinlichkeit dafür, daß die beobachtete Trajektorie sich stets unterhalb der Geraden $v = c$ befindet, zurückgeführt. Die Menge aller möglichen Trajektorien betrachten wir hier als Menge der Elementarereignisse. Eine Trajektorie ist durch $2m$ Verschiebungen bestimmt, wobei m Verschiebungen positiv und m negativ sind. Mit anderen Worten: Eine Trajektorie ist durch ein System von m Zahlen 1 und m Zahlen -1 bestimmt; folglich ist die Anzahl der Trajektorien gleich der Anzahl der Kombinationen von $2m$ Elementen zu je m, also gleich $\binom{2m}{m}$. Aus der Voraussetzung, daß die Zufallsvariablen $X_{11}, \ldots, X_{1m}, X_{21}, \ldots, X_{2m}$ unabhängig sind und dieselbe Verteilung haben, folgt, daß jede Trajektorie dieselbe Wahrscheinlichkeit hat. Diese Wahrscheinlichkeit ist gleich $1\Big/\binom{2m}{m}$. Um $Q^+_{m,m}(\lambda)$ zu bestimmen, müssen wir noch die Anzahl der Trajektorien

finden, die sich stets unterhalb der Geraden $v = c$ befinden. Zuerst wollen wir die Anzahl derjenigen Trajektorien bestimmen, die die Gerade $v = c$ schneiden. Zu diesem Zweck gehen wir folgendermaßen vor. Jeder Trajektorie, die die Gerade $v = c$ schneidet, ordnen wir umkehrbar eindeutig eine Trajektorie zu, die bis zum ersten Schnittpunkt mit der Geraden $v = c$ identisch mit der ersten ist und die, von diesem Schnittpunkt angefangen, die symmetrische Spiegelung der ursprünglichen Trajektorie an der Geraden $v = c$ ist. Den Verlauf der neuen Trajektorie rechts vom ersten Schnittpunkt mit der Geraden $v = 1$ stellt die punktierte Linie in Abb. 10.11.1 dar. Die neue Trajektorie endet im Punkt $(2m, 2c)$. Die Zahl der Verschiebungen nach oben beträgt $m + c$, die Anzahl der Verschiebungen nach unten $m - c$. Die Gesamtanzahl dieser Trajektorien ist $\binom{2m}{m+c} = \binom{2m}{m-c}$. Damit ist die Anzahl der Trajektorien, die nicht an die Gerade $v = c$ heranreichen, gleich $\binom{2m}{m} - \binom{2m}{m-c}$. Da alle Trajektorien mit derselben Wahrscheinlichkeit auftreten, erhalten wir für $0 < \lambda \leqq \sqrt{\frac{m}{2}}$ aus (10.11.14)

$$Q^+_{m,m}(\lambda) = \frac{\binom{2m}{m} - \binom{2m}{m-c}}{\binom{2m}{m}} = 1 - \frac{\binom{2m}{m-c}}{\binom{2m}{m}}. \tag{10.11.15}$$

Wenn wir noch die offenbar gültige Gleichung $Q^+_{m,m}(\lambda) = 0$ $(\lambda \leqq 0)$ beachten und schließlich die Tatsache, daß für $\lambda > \sqrt{\frac{m}{2}}$ auch $c > m$ ist, während doch keine Trajektorie die Gerade $v = m$ überschreiten kann, so daß also $Q^+_{m,m}(\lambda) = 1$ für $\lambda > \sqrt{\frac{m}{2}}$ gilt, so erhalten wir die Relation (10.11.10).

Wir gehen nun zum Beweis von (10.11.11) über. Zunächst bemerken wir, daß analog zur Gleichung (10.11.14) die Beziehung

$$\begin{aligned} Q_{m,m}(\lambda) &= P\left(\sqrt{\frac{m}{2}}\, D_{m,m} < \lambda\right) \\ &= P(m D_{m,m} < c) = P\left(\max_{1 \leqq k \leqq 2m} |V_k| < c\right) \end{aligned} \tag{10.11.16}$$

gilt. Auch hier führt die Berechnung der Verteilungsfunktion $Q_{m,m}(\lambda)$ für jedes natürliche c auf die Bestimmung der Wahrscheinlichkeit dafür, daß sich die beobachtete Trajektorie zwischen den Geraden $v = -c$ und $v = c$ befindet, ohne die beiden Geraden zu schneiden.

Wir bezeichnen die Gerade $v = c$ mit α, die Gerade $v = -c$ mit β und die Menge aller Trajektorien mit A. Wie schon bekannt, befinden sich in A genau $\binom{2m}{m}$ Elemente. Mit A_0 bezeichnen wir die Menge derjenigen Trajektorien, die zwischen den Geraden α und β liegen und sie nirgends schneiden. Mit A_1 bezeichnen wir die Menge derjenigen Trajektorien, die wenigstens einmal die Gerade α schneiden, mit A_2 die Menge aller der Trajektorien, die wenigstens einmal die Gerade α und dann die Gerade β schneiden, mit A_3 die Menge der Trajektorien, die wenigstens einmal die Geraden α und β in der Reihenfolge α, β, α schneiden. Allgemein bedeutet A_{2i-1} die Menge der Trajektorien, die wenigstens einmal die Geraden α und β in der Reihenfolge $\alpha, \beta, \ldots, \alpha$ schneiden (α tritt hier i-mal, β tritt $(i-1)$-mal auf, während A_{2i} die Menge derjenigen Trajektorien ist, die wenigstens einmal die Geraden α und β in der Reihenfolge $\alpha, \beta, \ldots, \alpha, \beta$ (α und β treten hier je i-mal auf) schneiden. Ähnlich bedeutet B_{2i-1} die Menge der Trajektorien, die wenigstens einmal die Geraden α und β in der Reihenfolge $\beta, \alpha, \ldots, \beta$ (β tritt i-mal, α tritt $(i-1)$-mal auf) schneiden, während B_{2i} die Menge derjenigen Trajektorien bezeichnet, die die Geraden α und β wenigstens einmal in der Reihenfolge $\beta, \alpha, \ldots, \beta, \alpha$ (β und α je i-mal) schneiden. Dann läßt sich leicht die Beziehung

$$A = A_0 + \sum_{i \geqq 1} [(A_{2i-1} - A_{2i}) + (B_{2i-1} - B_{2i})] \qquad (10.11.17)$$

beweisen. Dabei haben die Mengen in den eckigen Klammern für verschiedene i leeren Durchschnitt. Wir zeigten oben, daß die Anzahl der Elemente von A_1 (also auch von B_1) gleich $\binom{2m}{m-c}$ ist. Um die Anzahl der Elemente von A_2 zu finden, ordnen wir jeder zur Menge A_2 gehörigen Trajektorie in zwei Schritten umkehrbar eindeutig eine neue Trajektorie zu, die bis zum ersten Schnittpunkt mit der Geraden α identisch mit der ursprünglichen ist. Im ersten Schritt spiegeln wir die ursprüngliche Trajektorie von diesem Schnittpunkt, angefangen an der Geraden $v = c$. Die so entstehende Trajektorie verfolgen wir bis zum ersten Schnittpunkt mit der Geraden $v = 3c$. Von diesem Schnittpunkt angefangen, spiegeln wir die im ersten Schritt konstruierte Trajektorie an der Geraden $v = 3c$. Diese neue Trajektorie ordnen wir der ursprünglichen zu; sie endet im Punkt $(2m, 4c)$. Die Anzahl der Verschiebungen nach oben beträgt also $m + 2c$, die Anzahl der Verschiebungen nach unten $m - 2c$. Die Anzahl der Elemente von A_2 ist also gleich $\binom{2m}{m-2c}$. Dieselbe Anzahl von Elementen hat auch die Menge B_2

In Abb. 10.11.1 zeigt die gestrichelte Linie rechts vom Schnittpunkt mit der Geraden $v = 3$ den Verlauf der neuen Trajektorie.

Ähnlich kann man zeigen, daß die Anzahl der Elemente von A_j (also auch von B_j), wobei $j \leqq \left[\frac{m}{c}\right]$, gleich $\binom{2m}{m-jc}$ ist. Aus (10.11.17) erhalten wir, daß

die Anzahl der Elemente von A_0 gleich

$$\binom{2m}{m} - 2\sum_{i\geqq 1}\left[\binom{2m}{m-(2i-1)c} - \binom{2m}{m-2ic}\right] = \sum_{j=-[m/c]}^{[m/c]} (-1)^j \binom{2m}{m-jc}$$

ist.

Wenn wir berücksichtigen, daß alle Trajektorien mit derselben Wahrscheinlichkeit auftreten, erhalten wir aus (10.11.16) die Beziehung

$$Q_{m,m}(\lambda) = \sum_{j=-[m/c]}^{[m/c]} (-1)^j \frac{\binom{2m}{m-jc}}{\binom{2m}{m}} \qquad \left(\frac{1}{\sqrt{2m}} < \lambda \leqq \sqrt{\frac{m}{2}}\right). \tag{10.11.18}$$

Ist $\lambda \leqq \dfrac{1}{\sqrt{2m}}$, dann ist $c \leqq 1$. Da $\max\limits_{1\leqq k\leqq 2m} |V_k|$ größer als Null ist, erhalten wir $Q_{m,m}(\lambda) = 0$ für diese λ. Wenn dagegen $\lambda > \sqrt{\dfrac{m}{2}}$ ist, so ist $c > m$, und für solche λ ist $Q_{m,m}(\lambda) = 1$. Die Relation (10.11.11) wurde damit vollständig bewiesen.

D. Beweis des Smirnowschen Satzes. Wir gehen nun zum Beweis des Satzes 10.11.2 über und setzen dabei voraus, daß $n_1 = n_2 = m$ ist. Für $j = 1, 2, 3, \ldots$ haben wir

$$I_j = \binom{2m}{m-jc} : \binom{2m}{m} = \frac{(m!)^2}{(m-jc)!\,(m+jc)!}.$$

Wenn wir ähnlich wie in 10.5 schließen, erhalten wir bei Beachtung der Stirlingschen Formel $r! \approx \left(\dfrac{r}{e}\right)^r \sqrt{2\pi r}$ und der Näherungsgleichung $c = \lambda\sqrt{2m}$ die Formel

$$I_j \approx \left(1 - \frac{\lambda j\sqrt{2}}{\sqrt{m}}\right)^{j\lambda\sqrt{2m}-m} \left(1 + \frac{\lambda j\sqrt{2}}{\sqrt{m}}\right)^{-j\lambda\sqrt{2m}-m} \sqrt{\frac{m}{\left(\sqrt{m} - j\lambda\sqrt{2}\right)\left(\sqrt{m} + j\lambda\sqrt{2}\right)}}.$$

Daraus ergibt sich

$$\log I_j = -2j^2\lambda^2 + \sqrt{m}\,o\left(\frac{1}{\sqrt{m}}\right).$$

Schließlich erhalten wir

$$\lim_{m\to\infty} I_j = \exp(-2j^2\lambda^2). \tag{10.11.19}$$

Gehen wir in der Formel (10.11.10) mit $m \to \infty$ zur Grenze über und beachten wir dabei gleichzeitig die Formel (10.11.19) für $j = 1$, so erhalten wir (10.11.8).

Um (10.11.9) zu beweisen, bemerken wir, daß man zu jedem $\lambda > 0$ und jedem $\varepsilon > 0$ ein $j_0 > 0$ finden kann derart, daß

$$\exp(-2j_0^2\lambda^2) < \frac{\varepsilon}{16} \tag{10.11.20}$$

ist. Aus dem Leibnizschen Satz ergibt sich dann

$$\left|\sum_{|j|>j_0}(-1)^j \exp(-2j^2\lambda^2)\right| < 2\exp(-2(j_0+1)^2\lambda^2) < \frac{\varepsilon}{8}. \tag{10.11.21}$$

Wir setzen

$$R = \sum_{j_0<|j|\leq[m/c]} (-1)^j \frac{\binom{2m}{m-jc}}{\binom{2m}{m}}.$$

Da für $j > 0$ die Ungleichung $\binom{2m}{m-jc} > \binom{2m}{m-(j+1)c}$ und für $j < 0$ die Ungleichung $\binom{2m}{m-jc} > \binom{2m}{m-(j-1)c}$ gilt, gilt auch

$$|R| < 4\,\frac{\binom{2m}{m-j_0c}}{\binom{2m}{m}}.$$

Diese Ungleichung und die Formeln (10.11.19) und (10.11.20) führen für genügend große m zu

$$|R| < \frac{\varepsilon}{3}.$$

Dies und die Ungleichung (10.11.21) ziehen die Relation

$$\left|R - \sum_{|j|>j_0}(-1)^j \exp(-2\lambda^2 j^2)\right| < \frac{\varepsilon}{2} \tag{10.11.22}$$

nach sich.

Andererseits folgt aus (10.11.19), daß für genügend große m die Ungleichung

$$\left|\sum_{|j|\leq j_0}(-1)^j\left(\frac{\binom{2m}{m-jc}}{\binom{2m}{m}} - \exp(-2j^2\lambda^2)\right)\right| < \frac{\varepsilon}{2} \tag{10.11.23}$$

erfüllt ist. Gehen wir in (10.11.11) mit $m \to \infty$ zur Grenze über und benutzen wir die Ungleichungen (10.11.22) und (10.11.23), so erhalten wir (10.11.9).

Tafeln für die Verteilungsfunktionen $Q^+_{m,m}(\lambda)$ und $Q_{m,m}(\lambda)$ kann man für kleine m bei MASSEY [2] finden, ebenso bei SADOWSKI [1] und JANKO [1].

Es sei ausdrücklich betont, daß möglichst exakte Tafeln für $Q_{m,m}(\lambda)$ benutzt werden müssen, denn $Q_{m,m}(\lambda)$ kann, wie von HODGES [1] gezeigt wurde, selbst für ziemlich große m von der Grenzverteilung $Q(\lambda)$ bedeutend abweichen.

Beispiel 10.11.2. Wir entnehmen ein und derselben Gesamtheit, deren Merkmal X stetig verteilt sei, zwei unabhängige einfache Stichproben zu je 4 Elementen. Dabei beobachten wir die Größen

$$x_{11} = 1, \quad x_{12} = 3, \quad x_{13} = -7, \quad x_{14} = 2,$$
$$x_{21} = 6, \quad x_{22} = -4, \quad x_{23} = 0, \quad x_{24} = 8.$$

Wir haben hier

$$y_1 = -7, \quad y_2 = -4, \quad y_3 = 0, \quad y_4 = 1, \quad y_5 = 2, \quad y_6 = 3, \quad y_7 = 6, \quad y_8 = 8,$$

also

$$z_1 = 1, \quad z_2 = -1, \quad z_3 = -1, \quad z_4 = 1, \quad z_5 = 1, \quad z_6 = 1, \quad z_7 = -1, \quad z_8 = -1,$$
$$v_1 = 1, \quad v_2 = 0, \quad v_3 = -1, \quad v_4 = 0, \quad v_5 = 1, \quad v_6 = 2, \quad v_7 = 1, \quad v_8 = 0.$$

Die beobachteten Werte der Zufallsvariablen $D^+_{4,4}$ und $D_{4,4}$ betragen $d^+_{4,4} = d_{4,4} = \frac{1}{2}$. Wir suchen die Wahrscheinlichkeit dafür, daß wir nicht kleinere Werte als die beobachteten erhalten. Es gilt hier $\frac{\lambda}{\sqrt{2}} = \frac{1}{2}$, also $\lambda = \frac{1}{\sqrt{2}}$ und $c = 2$.

Aus (10.11.10) erhalten wir $Q^+_{4,4}\left(\frac{1}{\sqrt{2}}\right) = 0{,}6$, d. h., die Wahrscheinlichkeit dafür, daß $D^+_{4,4}$ nicht kleinere Werte als $\frac{1}{2}$ annimmt, ist gleich 0,4. Aus (10.11.11) erhalten wir $Q_{4,4}\left(\frac{1}{\sqrt{2}}\right) = 0{,}229$. Die gesuchte Ergänzung zur Eins beträgt also 0,771.

Beispiel 10.11.3. Wir entnehmen zwei einfache unabhängige Stichproben einer Gesamtheit, deren Merkmal X stetig verteilt sei. Der Umfang der ersten Stichprobe sei $n_1 = 70$, der der zweiten Stichprobe $n_2 = 100$. Wir beobachteten nun den Wert der Zufallsvariablen D_{n_1,n_2}. Er betrage $d_{n_1,n_2} = 0{,}18$. Wie groß ist die Wahrscheinlichkeit dafür, daß D_{n_1,n_2} kleiner als dieser beobachtete Wert ist?

Es gilt hier $n = \frac{n_1 n_2}{n_1 + n_2} = 41{,}2$, $\sqrt{n} = 6{,}42$, $\lambda = 0{,}18 \cdot 6{,}42 = 1{,}16$, also auf Grund des Smirnowschen Grenzwertsatzes

$$P(D_{n_1,n_2} < 0{,}18) = P\left(D_{n_1,n_2} < \frac{1{,}16}{6{,}42}\right) \approx Q(1{,}16).$$

In Tafel VIII finden wir $Q(1{,}16) \approx 0{,}8644$.

E. Wir geben noch einige weitere Informationen. Es sei

$$D_n^+ = \sup_{-\infty<x<\infty} \left(S_n(x) - F(x)\right),$$

$$D_{n_1,n_2}^- = \max_{-\infty<x<\infty} \left(S_{2n_2}(x) - S_{1n_2}(x)\right) = - \min_{-\infty<x<\infty} \left(S_{1n_1}(x) - S_{2n_2}(x)\right).$$

Einen Ausdruck für die genaue Verteilung der Stichprobenfunktion D_n^+ haben Wald und Wolfowitz [1] und später Birnbaum und Tingey [1] angegeben.

Unter der Voraussetzung, daß die Bedingungen des Satzes 10.11.1 erfüllt sind, hat N. W. Smirnow [1], [2] gezeigt, daß für $\lambda > 0$

$$\lim_{n\to\infty} P\left(D_n^+ \sqrt{n} < \lambda\right) = 1 - \exp(-2\lambda^2) \tag{10.11.24}$$

gilt.

In den zitierten Arbeiten hat Smirnow die Grenzverteilung des Zufallsvektors $(D_{n_1,n_2}^+, D_{n_1,n_2}^-)$ mit $n = \dfrac{n_1 n_2}{n_1 + n_2}$ angegeben.

Gnedenko und Rwatschewa [1] fanden für $n_1 = n_2 = m$ die exakte zweidimensionale Verteilung des Vektors $\left(\sqrt{\dfrac{1}{2}}\, m D_{m,m}^+, \sqrt{\dfrac{1}{2}}\, m D_{m,m}^-\right)$ und berechneten den Korrelationskoeffizienten zwischen $D_{m,m}^+$ und $D_{m,m}^-$

N. W. Smirnow [2] untersuchte die Grenzverteilung für die Anzahl der Schnittpunkte der empirischen Verteilungsfunktion $S_{1n_1}(x)$ mit der Kurve

$$S_{2n_2}(x) + \lambda \sqrt{\frac{n_1 + n_2}{n_1 n_2}}.$$

Michalewitsch [1] fand für $n_1 = n_2$ die exakte Verteilung für die Anzahl dieser Schnittpunkte.

Die Verteilung einer Stichprobenfunktion, die mit der Stichprobenfunktion D_n^+ eng verknüpft ist, untersuchten Birnbaum und Pyke [1] (siehe auch Aufgabe 10.14.19).

Kac [3] untersuchte die Verteilung der Stichprobenfunktion D_n unter der Voraussetzung, daß n eine Zufallsvariable mit Poissonverteilung ist.

Zu gewissen Ergebnissen in Richtung einer Verallgemeinerung des Satzes 10.11.1 von Kolmogoroff sowie der Beziehung (10.11.24) für Zufallsvektoren kamen Kiefer und Wolfowitz [3].

10.12. Der Satz von Rényi

Rényi [3] modifizierte den Satz 10.11.1 von Kolmogoroff folgendermaßen: Im Kolmogoroffschen Satz untersucht man die Differenzen $|S_n(x) - F(x)|$, ohne den Wert von $F(x)$ zu berücksichtigen. So kann zum Beispiel diese Differenz

im Punkt x den Wert 0,01 haben. Wenn in diesem Punkt $F(x) = 0{,}5$ ist, so beträgt die Differenz 2% des Wertes von $F(x)$. Ist in diesem Punkt aber $F(x) = 0{,}01$, so beträgt die Differenz 100% des Wertes von $F(x)$. Diese Möglichkeiten werden beim Kolmogoroffschen Satz nicht unterschieden. RÉNYI schlägt nun vor, anstelle der Differenz $S_n(x) - F(x)$ die Stichprobenfunktion $\dfrac{S_n(x) - F(x)}{F(x)}$ zu untersuchen. Die Voraussetzungen des Satzes 10.11.1 seien erfüllt. RÉNYI hat nun bewiesen, daß dann für beliebiges $a > 0$ die folgenden Beziehungen gelten:

$$\lim_{n\to\infty} P\left(\sqrt{n} \sup_{a \leq F(x)} \frac{S_n(x) - F(x)}{F(x)} < \lambda\right)$$
$$= \begin{cases} 0 & \text{für } \lambda \leqq 0, \\ \sqrt{\dfrac{2}{\pi}} \displaystyle\int_0^{\lambda\sqrt{a/(1-a)}} \exp\left(-\frac{1}{2}\lambda^2\right) d\lambda & \text{für } \lambda > 0, \end{cases} \tag{10.12.1}$$

$$\lim_{n\to\infty} P\left(\sqrt{n} \sup_{a \leq F(x)} \left|\frac{S_n(x) - F(x)}{F(x)}\right| < \lambda\right)$$
$$= \begin{cases} 0 & \text{für } \lambda \leqq 0, \\ \dfrac{4}{\pi} \displaystyle\sum_{j=0}^{\infty} (-1)^j \frac{\exp\left[-\dfrac{(2j+1)^2\pi^2}{8}\,\dfrac{1-a}{a\lambda^2}\right]}{2j+1} & \text{für } \lambda > 0. \end{cases} \tag{10.12.2}$$

Das Supremum auf der linken Seite dieser Formeln wählten wir für alle diejenigen Werte x, für welche $F(x) \geqq a$ mit beliebigem $a > 0$ ist.

Dieser Satz kann mit Nutzen in der Praxis angewandt werden.

Tafeln für die rechte Seite von (10.12.2) hat RÉNYI in der schon zitierten Arbeit [3] angegeben.

Die Verteilung der Stichprobenfunktion

$$\sqrt{n} \sup_{a \leq F(x)} \frac{S_n(x) - F(x)}{F(x)}$$

für endliche n hat ISHII [1] angegeben (siehe auch Aufgabe 10.14.22).

SCHMID [1] verallgemeinerte den Satz von RÉNYI für den Fall, daß die Verteilungsfunktion $F(x)$ Unstetigkeitsstellen aufweist.

WANG [1] hat für die empirischen Verteilungsfunktionen $S_{1n_1}(x)$ und $S_{2n_2}(x)$ von zwei unabhängigen einfachen Stichproben, die aus derselben Gesamtheit mit einer stetigen Verteilungsfunktion stammen, eine der Formel (10.12.1)

ähnliche Formel angegeben. Er bewies nämlich, daß die Relation

$$\lim_{n_2 \to \infty} P\left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \sup_{a \leqq S_{2n_2}(x)} \frac{S_{1n_1}(x) - S_{2n_2}(x)}{S_{2n_2}(x)} < \lambda\right)$$

$$= \begin{cases} 0 & \text{für} \quad \lambda \leqq 0, \\ \sqrt{\frac{2}{\pi}} \int\limits_0^{\lambda\sqrt{a/1-a}} \exp\left(-\frac{1}{2}\lambda^2\right) d\lambda & \text{für} \quad \lambda > 0 \end{cases}$$

für beliebiges $a > 0$ erfüllt ist, wenn für $n_1 \to \infty$ und $n_2 \to \infty$ die Relation

$$\frac{n_1}{n_2} \to d \leqq 1$$

gilt, wobei d eine gewisse Konstante ist.

Chang [1] fand die exakte Verteilung der Zufallsvariablen

$$\sup_{0 < F(x) \leqq a} \frac{S_n(x)}{F(x)}, \qquad \sup_{0 < S_n(x) \leqq a} \frac{S_n(x)}{F(x)},$$

wobei a konstant und $0 < a \leqq 1$ ist.

Manija [1] untersuchte die Verteilungen der Stichprobenfunktion

$$\sup_{a < x < b} |S_n(x) - F(x)|,$$

wobei a und b gewisse Konstanten sind. Kwit [1] untersuchte die Verteilung der Stichprobenfunktion

$$\sup_{-\infty < x \leqq a,\ b \leqq x < \infty} |S_n(x) - F(x)|.$$

Die Modifikationen der von Rényi, Chang, Manija und Kwit untersuchten Stichprobenfunktionen von Kolmogoroff und Smirnow sind Spezialfälle der folgenden von Anderson und Darling [1] untersuchten Stichprobenfunktionen:

$$D_n^* = \sup_{-\infty < x < \infty} \sqrt{n}\, |S_n(x) - F(x)|\, \psi[F(x)],$$

$$D_n^{**} = \sup_{-\infty < x < \infty} \sqrt{n}\, [S_n(x) - F(x)]\, \psi[F(x)];$$

hierbei bedeutet ψ eine gewisse Funktion.

Gewisse Modifikationen und Verallgemeinerungen der Kolmogoroffschen und Smirnowschen Stichprobenfunktionen wurden ferner in den Arbeiten von Rwatschewa [1], Tsao [1], Kuiper [2], Vincze [1], [2] sowie Dwass [1] behandelt.

Eine umfangreiche synthetische Beschreibung der Problematik um die Kolmogoroffschen und Smirnowschen Stichprobenfunktionen findet der Leser in der Arbeit von DARLING [1].

10.13. Das Problem mehrerer Stichproben

A. Der Satz 10.11.1 erlaubt es zu entscheiden, ob das Merkmal X in der Gesamtheit, welcher die betrachtete Stichprobe entnommen wurde, eine vorgegebene stetige Verteilungsfunktion $F(x)$ hat. Mit Hilfe des Satzes, den wir jetzt angeben, können wir entscheiden, ob das Merkmal X in k $(k \geqq 2)$ Gesamtheiten, aus denen die k Stichproben gezogen wurden, dieselbe vorgegebene stetige Verteilungsfunktion $F(x)$ besitzt..

Wir betrachten also k $(k \geqq 2)$ Gesamtheiten, in denen das Merkmal X dieselbe stetige Verteilungsfunktion $F(x)$ hat. Jeder dieser Gesamtheiten entnimmt man eine einfache Stichprobe vom Umfang n. Dabei seien die einzelnen Stichproben voneinander unabhängig. Anders ausgedrückt, wir haben es mit $n \cdot k$ unabhängigen Zufallsvariablen $X_{11}, \ldots, X_{1n}, X_{21}, \ldots, X_{2n}, \ldots, X_{k1}, \ldots, X_{kn}$ zu tun, die dieselbe Verteilung haben und in k Gruppen eingeteilt wurden; dabei liegt über jede dieser Zufallsvariablen eine Beobachtung vor. Wir bezeichnen mit $S_{nj}(x)$ die empirische Verteilungsfunktion der j-ten Stichprobe, d. h. derjenigen Stichprobe, die aus der j-ten $(j = 1, 2, \ldots, k)$ Gesamtheit stammt. Weiter setzen wir

$$\begin{aligned} D(n, j) &= \sqrt{n} \sup_{-\infty < x < \infty} |S_{nj}(x) - F(x)| \qquad (j = 1, 2, \ldots, k), \\ M_n &= \max_{1 \leqq j \leqq k} D(n, j). \end{aligned} \tag{10.13.1}$$

Die Zufallsvariablen $D(n, j)$ $(j = 1, 2, \ldots, k)$ sind unabhängig und haben dieselbe Verteilung.

Satz 10.13.1. *Es seien $S_{nj}(x)$ $(j = 1, \ldots, k)$ die empirischen Verteilungsfunktionen von k unabhängigen einfachen Stichproben aus Gesamtheiten, in denen das Merkmal X die gleiche stetige Verteilungsfunktion $F(x)$ hat. Dann gilt*

$$\lim_{n \to \infty} P(M_n < \lambda) = [Q(\lambda)]^k, \tag{10.13.2}$$

wobei $Q(\lambda)$ durch die Formel (10.11.2) *gegeben ist.*

Beweis. Im Sinne von Satz 10.11.1 bestehen die Relationen

$$\lim_{n \to \infty} P\big(D(n, j) < \lambda\big) = Q(\lambda) \qquad (j = 1, 2, \ldots, k). \tag{10.13.3}$$

Da das Ereignis $(M_n < \lambda)$ dem vereinten Ereignis $\big(D(n, 1) < \lambda, \ldots, D(n, k) < \lambda\big)$ gleichwertig ist und die Zufallsvariablen $D(n, j)$ $(j = 1, 2, \ldots, k)$ voneinander

unabhängig sind, erhalten wir aus (10.13.3)

$$\lim_{n\to\infty} P(M_n < \lambda) = \lim_{n\to\infty} P(D(n, j) < \lambda;\ j = 1, \ldots, k)$$
$$= \prod_{j=1}^{k} \lim_{n\to\infty} P(D(n, j) < \lambda) = [Q(\lambda)]^k.$$

Damit ist der Satz bewiesen.

Es ist klar, daß die Voraussetzung über die Gleichheit der Stichprobenumfänge unwesentlich ist. Wenn die Stichprobenumfänge nicht gleich sind und n_j der Umfang der j-ten ($j = 1, 2, \ldots, k$) Stichprobe ist, so bleibt der Satz richtig, falls alle n_j gegen Unendlich streben.

B. Wir bringen nun einen Satz, der eine Verallgemeinerung des Smirnowschen Satzes für k ($k \geqq 2$) Stichproben darstellt.

Wir betrachten k Stichproben mit den in 10.13.A genannten Eigenschaften; n_j und $S_{jn_j}(x)$ ($j = 1, 2, \ldots, k$) mögen die Gliederzahl bzw. die empirische Verteilungsfunktion der j-ten Stichprobe bedeuten. Wir setzen für $i = 1, 2, \ldots, k-1$

$$A_i^+(n_1, n_2, \ldots, n_k) = \max_{-\infty<x<\infty} \sum_{j=1}^{k} B_{in_j} \sqrt{n_j}\, S_{jn_j}(x), \tag{10.13.4}$$

$$A_i(n_1, n_2, \ldots, n_k) = \max_{-\infty<x<\infty} \left| \sum_{j=1}^{k} B_{in_j} \sqrt{n_j}\, S_{jn_j}(x) \right|, \tag{10.13.5}$$

wobei die B_{in_j} Konstanten sind (die von den Argumenten $n_1, n_2, \ldots, n_k, i, j$ abhängen).

Satz 10.13.2. *Es seien die $S_{jn_j}(x)$ ($j = 1, 2, \ldots, k$) empirische Verteilungsfunktionen von k einfachen Stichproben, die aus k Gesamtheiten entnommen wurden, in denen das Merkmal X die gleiche stetige Verteilungsfunktion hat. Ferner nehmen wir an, daß die Beziehungen*

$$\sum_{j=1}^{k} B_{in_j} \sqrt{n_j} = 0 \qquad (i = 1, 2, \ldots, k-1), \tag{10.13.6}$$

$$\lim_{n_1\to\infty, \ldots, n_k\to\infty} B_{in_j} = B_{ij} \qquad (i = 1, 2, \ldots, k-1;\ j = 1, 2, \ldots, k) \tag{10.13.7}$$

gelten, wobei

$$\sum_{j=1}^{k} B_{hj} B_{ij} = \begin{cases} 0 & \text{für} \quad h \neq i, \\ 1 & \text{für} \quad h = i \end{cases} \tag{10.13.8}$$

ist. Dann gelten für beliebige positive Zahlen $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}, \lambda$ die Beziehungen

$$\lim_{n_1\to\infty, \ldots, n_k\to\infty} P(A_i^+(n_1, n_2, \ldots, n_k) < \lambda_i,\ i = 1, 2, \ldots, k-1)$$
$$= \prod_{i=1}^{k-1} \left(1 - \exp(-2\lambda_i^2)\right), \tag{10.13.9}$$

$$\lim_{n_1\to\infty,\ldots,n_k\to\infty} P\big(A_i(n_1, n_2, \ldots, n_k) < \lambda_i,\ i = 1, 2, \ldots, k-1\big)$$
$$= \prod_{i=1}^{k-1} Q(\lambda_i), \tag{10.13.10}$$

wobei $Q(\lambda_i)$ *durch* (10.11.2) *gegeben ist, sowie*

$$\lim_{n_1\to\infty,\ldots,n_k\to\infty} P\Big(\max_{1\leqq i\leqq k-1} A_i^+(n_1, n_2, \ldots, n_k) < \lambda\Big)$$
$$= \big(1 - \exp(-2\lambda^2)\big)^{k-1}, \tag{10.13.11}$$

$$\lim_{n_1\to\infty,\ldots,n_k\to\infty} P\Big(\max_{1\leqq i\leqq k-1} A_i(n_1, n_2, \ldots, n_k) < \lambda\Big)$$
$$= \big(Q(\lambda)\big)^{k-1}. \tag{10.13.12}$$

Diesen Satz haben Fisz [4], Chang und Fisz [1] sowie Kiefer [1] bewiesen.

Es liegt auf der Hand, daß man verschiedene Koeffizienten B_{in_j} wählen kann, die den Gleichungen (10.13.6) bis (10.13.8) genügen. Setzen wir zusätzlich voraus, daß

$$\lim_{n_1\to\infty} \frac{n_j}{n_1} = a_j > 0 \qquad (j = 2, \ldots, k)$$

gilt, so können wir das Koeffizientensystem

$$B_{in_j} = \begin{cases} -\sqrt{n_j n_{i+1}/N_i N_{i+1}} & \text{für } j = 1, 2, \ldots, i, \\ \sqrt{N_i/N_{i+1}} & \text{für } j = i+1, \\ 0 & \text{für } j = i+2, \ldots, k \end{cases} \tag{10.13.13}$$

betrachten, wobei $N_r = n_1 + n_2 + \cdots + n_r$ ist. Wir können dann

$$A_i^+(n_1, n_2, \ldots, n_k) = \sqrt{n_{i+1} N_i/N_{i+1}}\, D_i^+(N_i, n_{i+1}),$$

$$A_i(n_1, n_2, \ldots, n_k) = \sqrt{n_{i+1} N_i/N_{i+1}}\, D_i(N_i, n_{i+1})$$

schreiben, wobei

$$\begin{aligned} D_i^+(N_i, n_{i+1}) &= \max_x \Big[S_{i+1, n_{i+1}}(x) - \frac{1}{N_i}\sum_{j=1}^{i} n_j S_{in_j}(x)\Big], \\ D_i(N_i, n_{i+1}) &= \max_x \Big|S_{i+1, n_{i+1}}(x) - \frac{1}{N_i}\sum_{j=1}^{i} n_j S_{jn_j}(x)\Big| \end{aligned} \tag{10.13.14}$$

gilt. Für $k = 4$ haben wir z. B. die folgenden drei Zufallsvariablen:

$$D_1(N_1, n_2) = \max_x |S_{2n_2}(x) - S_{1n_1}(x)|,$$

$$D_2(N_2, n_3) = \max_x \left| S_{3n_3}(x) - \frac{n_1 S_{1n_1}(x) + n_2 S_{2n_2}(x)}{n_1 + n_2} \right|,$$

$$D_3(N_3, n_4) = \max_i \left| S_{4n_4}(x) - \frac{n_1 S_{1n_1}(x) + n_2 S_{2n_2}(x) + n_3 S_{3n_3}(x)}{n_1 + n_2 + n_3} \right|.$$

Wie CHANG und FISZ [2] sowie KIEFER [1] gezeigt haben, sind die Zufallsvariablen $D_1^+(N_1, n_2), \ldots, D_{k-1}^+(N_{k-1}, n_k)$ unabhängig. Das gleiche gilt für die Zufallsvariablen $D_1(N_1, n_2), \ldots, D_{k-1}(N_{k-1}, n_k)$.

GICHMAN [5] und KIEFER [1] haben eine andere Verallgemeinerung des Smirnowschen Satzes für k Stichproben gegeben, und zwar bewiesen sie den folgenden Satz:

Satz 10.13.3. *Es seien die $S_{jn_j}(x)$ $(j = 1, 2, \ldots, k)$ empirische Verteilungsfunktionen von k einfachen Stichproben, die aus k Gesamtheiten entnommen wurden, in denen das Merkmal X die gleiche stetige Verteilungsfunktion hat. Wir setzen*

$$S_{n_1,\ldots,n_k}(x) = \frac{1}{N_k} \sum_{j=1}^{k} n_j S_{jn_j}(x),$$

$$D^2_{n_1,\ldots,n_k} = \max_{-\infty<x<\infty} \sum_{j=1}^{k} n_j \left(S_{jn_j}(x) - S_{n_1,\ldots,n_k}(x)\right)^2.$$

Dann gilt

$$\lim_{n_1\to\infty,\ldots,n_k\to\infty} P(D_{n_1,\ldots,n_k} < \lambda)$$

$$= \frac{4}{\Gamma\left(\frac{1}{2}(k-1)\right)(2\lambda^2)^{(k-1)/2}} \sum_{s=1}^{\infty} \frac{\mu_s^{k-3}}{\left(J_{(k-3)/2}(\mu_s)\right)^2} \exp\left(-\frac{\mu_s^2}{2\lambda^2}\right), \qquad (10.13.15)$$

wobei μ_s die positive s-te Wurzel der Besselschen Funktion $J_{(k-3)/2}(z)$ ist.

Sämtliche Sätze aus 10.13.B haben wir ohne Beweis angegeben, da die Wiedergabe der Beweise über die Grenzen dieses Buches hinaus geht. Der Grundgedanke der Beweise ist der Grundidee zum Beweis des Satzes von KOLMOGOROFF ähnlich, den wir in 10.11.C gegeben hatten.

Andere Verallgemeinerungen des Smirnowschen Satzes für den Fall von drei Stichproben, die mit elementaren Methoden hergeleitet wurden, findet der Leser in den Arbeiten von OZOLS [1] (vgl. Aufgabe 10.14.20) und H. T. DAVID [1], und für eine beliebige endliche Anzahl von Stichproben in der Arbeit von BIRNBAUM und HALL [1] (siehe auch DWASS [2]).

10.14. Aufgaben und Ergänzungen

1. Einer Grundgesamtheit, in der das Merkmal X die Normalverteilung $N(0; 1)$ hat, wurde eine einfache zehngliedrige Stichprobe entnommen. Dabei erhielt man die Werte $x_1 = 0$, $x_2 = 0{,}2$, $x_3 = 0{,}25$, $x_4 = -0{,}3$, $x_5 = -0{,}1$, $x_5 = 2$, $x_7 = 0{,}15$, $x_8 = 1$, $x_9 = -0{,}7$, $x_{10} = -1$.

   a) Man bestimme die empirische Verteilungsfunktion $S_n(x)$ dieser Stichprobe.

   b) Man berechne $E\big(S_n(x)\big)$ und $D^2\big(S_n(x)\big)$ an der Stelle $x = 0{,}2$.

   c) Man berechne $\Phi_{kn}(x)$ für $k = 6$, $n = 10$ und $x = 0{,}2$.

2. Es sei $\zeta_k^{(n)}$ Positionsfunktion einer aus einer Gesamtheit entnommenen einfachen Stichprobe, in der die Eigenschaft X die stetige Verteilungsfunktion $F(x)$ hat. Man zeige, daß die Zufallsvariable $F(\zeta_k^{(n)})$ die Betaverteilung mit den Parametern k und $n - k + 1$ hat.

3. Man beweise, daß man die Behauptung des Satzes 10.4.1 unter den gleichen Voraussetzungen durch eine schärfere Behauptung ersetzen kann, und zwar, daß die Folge $\{\zeta_k^{(n)}\}$ für $n \to \infty$ mit der Wahrscheinlichkeit 1 gegen a_λ konvergiert.

   Hinweis. Man beweise, daß für jedes $\varepsilon > 0$
   $$\sum_{n=1}^{\infty} P(|\zeta_k^{(n)} - a_\lambda| > \varepsilon) < \infty$$
   gilt, und wende dann das Lemma von Borel-Cantelli an.

4. Es sei $\Phi_{kn}(x)$ durch die Formel (10.3.6) definiert, und es sei k konstant. Man zeige, daß für jedes x die Ungleichung
   $$\frac{1 - a_n}{(k-1)!} \int_0^{nF(x)} e^{-x} x^{k-1}\, dx \leqq \Phi_{kn}(x) \leqq \frac{1 + b_n}{(k-1)!} \int_0^{nF(x)} e^{-x} x^{k-1}\, dx$$
   gilt; $F(x)$ ist die theoretische Verteilungsfunktion, $a_n > 0$, $b_n > 0$, $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n = 0$.

5. Es sei $\{\zeta_k^{(n)}\}$ eine Folge von Positionsstichprobenfunktionen bei konstantem k, und es sei $F(x)$ die theoretische Verteilungsfunktion.

   a) Man zeige, daß die Folge $\{\zeta_k^{(n)} - A_k^{(n)}\}$, wobei die $A_k^{(n)}$ gewisse Konstanten sind, dann und nur dann stochastisch gegen Null konvergiert, wenn für jedes $\varepsilon > 0$ die folgenden Beziehungen erfüllt sind:
   $$\lim_{n\to\infty} nF(A_k^{(n)} - \varepsilon) = 0, \qquad \lim_{n\to\infty} nF(A_k^{(n)} + \varepsilon) = \infty.$$
   Hinweis. Man verwende die Aufgabe 4.

   b) Man löse eine analoge Aufgabe für $k = n - l + 1$ bei konstantem l.

6. (Fortsetzung) a) Man zeige, daß die Folge $\{\zeta_k^{(n)} - x_0\}$, wobei k konstant ist, für $n \to \infty$ stochastisch gegen Null konvergiert, wenn für ein gewisses x_0 stets $F(x) = 0$ für $x \leqq x_0$ und $F(x) > 0$ für $x > x_0$ gilt.

   b) Man formuliere und beweise ein analoges Ergebnis, wenn für ein gewisses x_0 stets $F(x) < 1$ für $x < x_0$ und $F(x) = 1$ für $x \geqq x_0$ gilt.

   Hinweis. Man stütze sich auf die Aufgabe 5.

7. Man beweise:

a) Gilt $F(x) > 0$ für jedes x, so ist die Folge $\{\zeta_k^{(n)} - A_k^{(n)}\}$, wobei k konstant ist, für eine gewisse Konstantenfolge $\{A_k^{(n)}\}$ für $n \to \infty$ dann und nur dann stochastisch gegen Null konvergent, wenn für jedes $\varepsilon > 0$ die folgende Relation erfüllt ist:

$$\lim_{x \to -\infty} \frac{F(x-\varepsilon)}{F(x)} = 0.$$

b) Gilt $F(x) < 1$ für jedes x, so konvergiert die Folge $\{\zeta_k^{(n)} - A_k^{(n)}\}$, $k = n - l + 1$ bei konstantem l für $n \to \infty$ dann und nur dann stochastisch gegen Null, wenn für jedes $\varepsilon > 0$ die Beziehung

$$\lim_{x \to \infty} \frac{1 - F(x+\varepsilon)}{1 - F(x)} = 0$$

erfüllt ist (für $k = 1$ oder $k = n$ siehe GNEDENKO [8], für den allgemeinen Fall N. W. SMIRNOW [3]).

8. Es seien $F(x)$ bzw. a_λ $(0 < \lambda < 1)$ die Verteilungsfunktion bzw. das Quantil der Zufallsvariablen X. Ferner sei $\zeta_k^{(n)}$ das Quantil einer n-gliedrigen einfachen Stichprobe. Man zeige: Soll für gewisse Folgen von Konstanten $a_n > 0$ und b_n die Beziehung

$$\lim_{n \to \infty} P\left(\frac{\zeta_k^{(n)} - b_n}{a_n} < z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz$$

gelten, so ist notwendig und hinreichend (N. W. SMIRNOW [3]), daß für jedes x die Beziehung

$$\lim_{n \to \infty} \frac{F(a_n x + b_n) - \lambda}{\sqrt{\lambda(1-\lambda)}} \sqrt{n} = x$$

erfüllt ist.

9. Es seien $F(x)$, $f(x)$, a_{λ_1} bzw. a_{λ_2} $(0 < \lambda_1 < \lambda_2 < 1)$ die Verteilungsfunktion, die Dichtefunktion bzw. Quantile der Zufallsvariablen X. Ferner seien $\zeta_{k_1}^{(n)}$ und $\zeta_{k_2}^{(n)}$ $(k_1 = [n\lambda_1] + 1$, $k_2 = [n\lambda_2] + 1)$ Quantile einer einfachen Stichprobe. Man zeige: Ist die Dichtefunktion $f(x)$ stetig und positiv an den Stellen $x = a_{\lambda_1}$ und $x = a_{\lambda_2}$, so ist der Zufallsvektor $(\zeta_{k_1}^{(n)}, \zeta_{k_2}^{(n)})$ für $n \to \infty$ asymptotisch normal verteilt mit den Parametern

$$m_1 = a_{\lambda_1}, \quad m_2 = a_{\lambda_2},$$

$$\sigma_1^2 = \frac{\lambda_1(1-\lambda_1)}{n\left(f(a_{\lambda_1})\right)^2}, \quad \sigma_2^2 = \frac{\lambda_2(1-\lambda_2)}{n\left(f(a_{\lambda_2})\right)^2}, \quad \varrho = \sqrt{\frac{\lambda_1(1-\lambda_2)}{\lambda_2(1-\lambda_1)}}.$$

Man verallgemeinere dieses Ergebnis für r $(r > 2)$ Positionsstichprobenfunktionen.

10. Es sei $\{\zeta_k^{(n)}\}$ eine Folge von Stichprobenfunktionen bei konstantem k, und es sei $F(x)$ die Verteilungsfunktion der Zufallsvariablen X in einer Grundgesamtheit. Man beweise: Soll für gewisse Folgen von Konstanten $a_n > 0$ und b_n die Beziehung

$$\lim_{n \to \infty} P\left(\frac{\zeta_k^{(n)} - b_n}{a_n} < z\right) = \begin{cases} 0 & \text{für } z \leqq 0, \\ \dfrac{1}{(k-1)!} \displaystyle\int_0^z e^{-z} z^{k-1}\, dz & \text{für } z > 0 \end{cases}$$

gelten, so ist notwendig und hinreichend, daß

$$\lim_{n\to\infty} nF(a_n x + b_n) = \begin{cases} 0 & \text{für} \quad x \leqq 0, \\ x & \text{für} \quad x > 0 \end{cases}$$

ist (GNEDENKO [8], N. W. SMIRNOW [3]).

11. Man zeige, daß die Zufallsvariablen $Z_k^{(n)}$ und $U_k^{(n)}$ (siehe 10.6) für $n \to \infty$ asymptotisch unabhängig sind.

12. Es seien $\zeta_1^{(n)}, \zeta_2^{(n)}, \ldots, \zeta_n^{(n)}$ Positionsfunktionen einer einfachen Stichprobe, die aus einer Gesamtheit herausgegriffen wurde, in der X im Intervall $[0, 1]$ gleichverteilt ist. Man beweise, daß die Zufallsvariablen $V_1, V_2, \ldots, V_n$, wobei

$$V_1 = \frac{\zeta_1^{(n)}}{\zeta_2^{(n)}}, \ldots, V_{n-1} = \frac{\zeta_{n-1}^{(n)}}{\zeta_n^{(n)}}, \quad V_n = \zeta_n^{(n)}$$

gesetzt ist, unabhängig sind und daß darüber hinaus V_j $(j = 1, 2, \ldots, n-1)$ im Intervall $[0, 1]$ gleichverteilt ist (MALMQUIST [1]).

13. Es seien $\zeta_1^{(2)}$ und $\zeta_2^{(2)}$ Positionsfunktionen einer einfachen Stichprobe, die aus einer Gesamtheit entnommen wurde, in der das Merkmal X die Verteilungsfunktion $F(x)$ und die Dichtefunktion $f(x)$ besitzt und $F(0) = 0$ ist. Man zeige: Eine notwendige und hinreichende Bedingung dafür, daß $f(x)$ die Form

$$f(x) = \begin{cases} \dfrac{m+1}{b^{m+1}} x^m & \text{für} \quad 0 \leqq x \leqq b, \\ 0 & \text{für} \quad x < 0,\ x > b \end{cases} \qquad (*)$$

hat, ist die Unabhängigkeit der Zufallsvariablen V_1 und V_2, wobei

$$V_1 = \frac{\zeta_1^{(2)}}{\zeta_2^{(2)}}, \qquad V_2 = \zeta_2^{(2)}$$

gesetzt ist. Wenn darüber hinaus in der Formel (*) für die Dichtefunktion $f(x)$ noch $b = 1$ gesetzt ist, so hat V_1 auch eine Verteilung, die durch dieselbe Dichtefunktion ausgedrückt ist (FISZ).

14. Man zeige (THOMPSON [1]): Ist $F(x)$ eine stetige Verteilungsfunktion und ist $a_{1/2}$ deren Mediane, so gilt

$$P(\zeta_k^{(n)} < a_{1/2} < \zeta_{n-k+1}^{(n)}) = 1 - \frac{2}{B(k, n-k+1)} \int_0^{0,5} t^{k-1}(1-t)^{n-k}\, dt.$$

Hinweis. Man beachte, daß $F(x)$ im Intervall $[0, 1]$ gleichverteilt ist.

15. Man zeige, daß die Dichtefunktion $h(w)$ der Breite W einer einfachen n-gliedrigen Stichprobe, die aus einer Gesamtheit entnommen wurde, in der das Merkmal X die Dichtefunktion $f(x)$ hat, die folgende Form besitzt:

$$h(w) = \begin{cases} n(n-1) \displaystyle\int_{-\infty}^{\infty} f(u) f(w+u) [F(w+u) - F(u)]^{n-2}\, du & \text{für} \quad 0 \leqq w < \infty, \\ 0 & \text{für} \quad w < 0. \end{cases}$$

16. a) Man beweise: Ist $F(x)$ eine stetige Verteilungsfunkiton, so hat die Stichprobenfunktion $W = P(L_1 \leqq X \leqq L_2)$ mit $L_1 = \zeta_k^{(n)}$ und $L_2 = \zeta_{n-k+1}^{(n)}$ $(k < n - k + 1)$ die Dichtefunktion

$$h(w) = \begin{cases} \dfrac{n!}{(n-2k)!\,(2k-1)!}\, w^{n-2k}(1-w)^{2k-1} & \text{für } 0 \leqq w \leqq 1, \\ 0 & \text{für } w < 0,\ w > 1. \end{cases}$$

b) Man bestimme die Momente $E(W)$ und $D^2(W)$.

c) Man leite die Formel für $h(w)$ her, wobei $L_1 = \zeta_k^{(n)}$ und $L_2 = \zeta_{n-l+1}^{(n)}$ $(k < n - l + 1)$ ist.

17. Es seien die Bedingungen des Satzes 10.11.2 erfüllt, und es sei $n_1 = n_2 = m$. Ferner bedeute C_m die Anzahl der Punkte x_k, für die $S_{1m}(x_k) \geqq S_{2m}(x_k + 0)$ gilt. Man zeige, daß dann (Chung und Feller [1], Gnedenko und Michalewitsch [1])

$$P(C_m = k) = \frac{1}{m+1} \qquad (k = 0, 1, \ldots, m)$$

ist.

Hinweis. Man wende ähnliche Überlegungen an wie beim Beweis des Lemmas in 10.11.

18. Es seien die Bedingungen des Satzes 10.11.2 erfüllt, und es sei $n_1 = n_2 = m$. Ferner seien $y_1 < y_2 < \cdots < y_{2m}$ die in aufsteigender Reihenfolge geordneten Werte der beiden Stichproben. Dann konvergiert die Wahrscheinlichkeit dafür, daß für alle zwischen y_{k_1} und y_{k_2} eingeschlossenen x-Werte die Ungleichung $S_{1m}(x) \leqq S_{2m}(x)$ gilt, für $m \to \infty$ gegen

$$\frac{1}{\pi} \arcsin \sqrt{\frac{\alpha(1-\beta)}{\beta(1-\alpha)}} \qquad \left(\alpha = \lim_{m\to\infty} \frac{k_1}{2m},\quad \beta = \lim_{m\to\infty} \frac{k_2}{2m}\right).$$

Das ist die *Arcussinusregel* (Lévy [7], Erdös und Kac [2], Gnedenko [7], Gichman [1]).

19. Es sei $(x_1, x_2, \ldots, x_n)$ eine einfache Stichprobe aus einer Gesamtheit, in der $F(x) = x$ $(0 \leqq x \leqq 1)$ ist, und es seien $y_1 < y_2 < \cdots < y_n$ die geordneten Werte von x. Ferner sei y^* der Wert y_k, für den $D_n^+ = \sup\,[S_n(x) - x]$ den maximalen Wert annimmt, und k^* bezeichne den Index k für $y_k = y^*$. Y^* und K^* seien Zufallsvariable, deren mögliche Werte y^* bzw. k^* sind. Man beweise die folgenden Behauptungen (Birnbaum und Pyke [1]; siehe auch Kuiper [1]):

a) Y^* ist im Intervall $[0, 1]$ gleichverteilt.

b) $P(K^* = k) = n^{-n} \sum\limits_{i=n-k}^{n-1} \dfrac{1}{i+1} \dbinom{n}{i} i^i (n-i)^{n-i-1}$.

c) Wie kann man dieses Ergebnis formulieren, wenn $F(x)$ eine beliebige stetige Verteilungsfunktion ist?

20. Es seien $S_{1n}(x)$, $S_{2n}(x)$ und $S_{3n}(x)$ empirische Verteilungsfunktionen einfacher Stichproben aus einer Gesamtheit, in der X eine stetige Verteilungsfunktion hat. Man zeige,

daß dann die folgende Beziehung für jedes $\lambda > 0$ gilt (Ozols [1]):

$$\lim_{n\to\infty} P\left(\sqrt{n/2}\max_x [S_{1n}(x) - S_{2n}(x)] < \lambda, \quad \sqrt{n/2}\max_x [S_{2n}(x) - S_{3n}(x)] < \lambda\right)$$
$$= 1 - \exp(-2\lambda^2) + 2\exp(-6\lambda^2) - \exp(-8\lambda^2).$$

Hinweis. Man stelle ähnliche Überlegungen an wie im Beweis des Lemmas in 10.11.C.

21. Man finde zwei verschiedene Koeffizientensysteme B_{in_j}, die die Beziehungen (10.13.6) bis (10.13.8) erfüllen.

22. (Bezeichnungen von 10.12.) Man beweise die Gültigkeit der Beziehung

$$\lim_{n\to\infty} P\left(\sqrt{n}\sup_{0\leq F(x)\leq 1}\frac{S_n(x) - F(x)}{F(x)} < \lambda\right) = \begin{cases} 0 & \text{für } \lambda < 0, \\ \dfrac{\lambda}{1+\lambda} & \text{für } \lambda \geqq 0. \end{cases}$$

(Dieser Satz wurde erstmalig von Daniels [2] bewiesen; später wurde er unabhängig voneinander von Hannan, Robbins, Chapman und Dempster entdeckt. In der Arbeit von Dempster [1] ist der Beweis angegeben.)

11. ABRISS DER ITERATIONSTHEORIE

11.1. Einleitende Bemerkungen

Die Iterationstheorie erlaubt, aus der Reihenfolge, in der die einzelnen Stichprobenelemente auftreten, Schlüsse zu ziehen. Befinden sich zum Beispiel in einer Stichprobe fehlerlose und fehlerhafte Stücke, so haben alle bis jetzt angegebenen Stichprobenfunktionen nur die *Anzahl* der fehlerlosen bzw. fehlerhaften Stichprobenelemente berücksichtigt, während die *Reihenfolge* ihres Auftretens nicht beachtet wurde. Dagegen berücksichtigt die Iterationstheorie die Reihenfolge, in der die einzelnen Stichprobenelemente auftreten, und erlaubt deshalb, die Information, die uns die Stichprobe vermittelt, besser auszunutzen.

11.2. Definition der Iterationen

Es seien X_j ($j = 1, 2, \ldots, n$) unabhängige Zufallsvariable mit derselben Zweipunktverteilung:

$$P(X_j = a_1) = p, \qquad P(X_j = a_2) = 1 - p.$$

Wir betrachten eine Stichprobe $(x_1, x_2, \ldots, x_n)$. Jede Stichprobe besteht hier aus Elementen a_1 und a_2, die sich wiederholen können und in einer bestimmten Reihenfolge auftreten.

Definition 11.2.1. Die Folge $(x_j, \ldots, x_{j+l})$, bei welcher $j = 1, 2, \ldots, n$ und $l = 0, 1, \ldots, n - j$ sein kann, wird eine *Iteration* genannt, wenn

$$x_{j-1} \neq x_j = \cdots = x_{j+l} \neq x_{j+l+1} \tag{11.2.1}$$

ist. Die Zahl $l + 1$ nennt man die *Länge der Iteration.*

Für $j = 1$ ist „$x_{j-1} \neq$“ auf der linken Seite von (11.2.1) überflüssig, während „$\neq x_{j+l+1}$“ für $j + l = n$ auf der rechten Seite von (11.2.1) überflüssig ist.

Mit K_{ij} wollen wir die Anzahl aller derjenigen Iterationen bezeichnen, die in einer Stichprobe vorkommen, die Länge j haben und nur aus a_i bestehen, wobei $i = 1$ oder 2 ist. Unter N_i $(i = 1, 2)$ verstehen wir die Anzahl aller a_i, die in der Stichprobe vorkommen. Es ist also $N_1 + N_2 = n$. Hier gelten die Formeln

$$\sum_j j K_{ij} = N_i \qquad (i = 1, 2). \tag{11.2.2}$$

Endlich verstehen wir unter K_i $(i = 1, 2)$ die Anzahl aller derjenigen Iterationen, die nur aus Gliedern a_i bestehen, während mit K die Gesamtanzahl aller Iterationen bezeichnet sei. Dann ist

$$K_i = \sum_j K_{ij} \qquad (i = 1, 2), \tag{11.2.3}$$

$$K = K_1 + K_2.$$

Beispiel 11.2.1. Wir betrachten die Stichprobe

$$a_1, a_1, a_2, a_2, a_2, a_1, a_2, a_1, a_1, a_1, a_1, a_2, a_2, a_2.$$

Hier haben wir eine a_1-Iteration der Länge 1, eine a_2-Iteration der Länge 1, eine a_1-Iteration der Länge 2, zwei a_2-Iterationen der Länge 3 und endlich eine a_1-Iteration der Länge 4. Für diese Stichprobe sind die beobachteten Werte k_{ij} der Zufallsvariablen K_{ij} gleich

$$k_{11} = k_{21} = k_{12} = k_{14} = 1, \quad k_{23} = 2.$$

Weiter gilt $n_1 = n_2 = 7$. Die beobachteten Werte k_1, k_2 und k der Zufallsvariablen K_1, K_2 und K betragen

$$k_1 = k_2 = 3, \quad k = 6.$$

11.3. Die Verteilungen der Iterationsanzahlen

A. Unsere Aufgabe besteht nun darin, die Verteilungen der oben eingeführten Zufallsvariablen zu finden. Zunächst bemerken wir, daß K_1 und K_2 sich höchstens um eine Einheit voneinander unterscheiden können; denn es sind für die beobachteten Werte der Zufallsvariablen K_1 und K_2 nur die folgenden drei Fälle möglich:

$$1.\ k_1 = k_2, \qquad 2.\ k_1 = k_2 + 1, \qquad 3.\ k_1 = k_2 - 1.$$

Wir führen die Funktion $G(k_1, k_2)$ ein, die folgendermaßen definiert sei:

$$G(k_1, k_2) = \begin{cases} 0 & \text{für} \quad |k_1 - k_2| > 1, \\ 1 & \text{für} \quad |k_1 - k_2| = 1, \\ 2 & \text{für} \quad k_1 = k_2. \end{cases} \tag{11.3.1}$$

Die Funktion $G(k_1, k_2)$ gibt uns an, auf wieviel verschiedene Arten man k_1 Iterationen von a_1 allein und k_2 Iterationen von a_2 allein erhalten kann. Man kann nämlich für $k_1 = k_2$ auf zweierlei Arten k_1 Iterationen von a_1 und k_2 Iterationen von a_2 erhalten, während für $k_1 \neq k_2$ (also für $|k_1 - k_2| = 1$) man nur auf eine einzige Art diese Iterationen erhalten kann.

Wir bezeichnen mit $P(k_{1j}, k_{2j}, n_1, n_2)$ die Wahrscheinlichkeitsfunktion der Zufallsvariablen

$$(K_{11}, \ldots, K_{1n_1}, K_{21}, \ldots, K_{2n_2}, N_1, N_2).$$

Es gilt:

$$\begin{aligned} &P(k_{1j}, k_{2j}, n_1, n_2) \qquad (11.3.2)\\ &= P(K_{11} = k_{11}, \ldots, K_{1n_1} = k_{1n_1}, K_{21} = k_{21}, \ldots, K_{2n_2} = k_{2n_2}, N_1 = n_1, N_2 = n_2)\\ &= P(K_{11} = k_{11}, \ldots, K_{1n_1} = k_{1n_1}, K_{21} = k_{21}, \ldots, K_{2n_2} = k_{2n_2} \mid N_1 = n_1, N_2 = n_2)\\ &\quad \times P(N_1 = n_1, N_2 = n_2). \end{aligned}$$

Aus der Unabhängigkeit der Zufallsvariablen X_k und daraus, daß ihre Verteilung dieselbe ist, folgt: Für gegebene n_1 und n_2 ist jede Reihenfolge der Glieder a_1 und a_2 gleichwahrscheinlich. Weiter bemerken wir, daß man die k_i $(i = 1, 2)$ Iterationen der Glieder a_i, wobei k_{i1} Iterationen die Länge 1, k_{i2} Iterationen die Länge 2, ..., k_{in_i} Iterationen die Länge n_i haben, auf

$$\frac{k_i!}{k_{i1}! \cdots k_{in_i}!}$$

verschiedene Arten anordnen kann. Berücksichtigen wir noch die Definition der Funktion $G(k_1, k_2)$, dann finden wir, daß wir k_1 aus Gliedern a_1 bestehende Iterationen, wobei k_{11} Iterationen die Länge 1, ..., k_{1n_1} die Länge n_1 haben, sowie k_2 Iterationen, gebildet aus Gliedern a_2, wobei k_{21} Iterationen die Länge 1, ..., k_{2n_2} die Länge n_2 haben, auf

$$\frac{k_1!}{k_{11}! \cdots k_{1n_1}!} \frac{k_2!}{k_{21}! \cdots k_{2n_2}!} G(k_1, k_2)$$

verschiedene Arten anordnen können. Da man die n Glieder, unter denen sich n_1 Glieder a_1 und $n_2 = n - n_1$ Glieder a_2 befinden, auf $\binom{n}{n_1}$ verschiedene Arten erhalten kann, ist die bedingte Wahrscheinlichkeit, die auf der rechten Seite der Formel (11.3.2) vorkommt, gleich

$$\frac{k_1!}{k_{11}! \cdots k_{1n_1}!} \frac{k_2!}{k_{21}! \cdots k_{2n_2}!} G(k_1, k_2) \frac{n_1! n_2!}{n!}.$$

Berücksichtigen wir schließlich noch

$$P(N_1 = n_1, N_2 = n_2) = \frac{n!}{n_1! n_2!} p^{n_1}(1 - p)^{n_2},$$

so erhalten wir endgültig

$$P(k_{1j}, k_{2j}, n_1, n_2) = \frac{k_1!}{k_{11}! \cdots k_{1n_1}!} \frac{k_2!}{k_{21}! \cdots k_{2n_2}!} G(k_1, k_2) p^{n_1} (1-p)^{n_2}. \tag{11.3.3}$$

B. Wir wollen jetzt die Randwahrscheinlichkeitsfunktionen $P(k_{1j}, n_1, n_2)$ der Zufallsvariablen $(K_{11}, \ldots, K_{1n_1}, N_1, N_2)$ bestimmen.

Diese Bestimmung führen wir in zwei Schritten durch. Zuerst wollen wir die Wahrscheinlichkeitsfunktion $P(k_{1j}, k_2, n_1, n_2)$ der Zufallsvariablen $(K_{11}, \ldots, K_{1n_1}, K_2, N_1, N_2)$ und dann die gesuchte Wahrscheinlichkeitsfunktion finden.

Wir berechnen für feste n_2 und k_2 den Ausdruck

$$\sum \frac{k_2!}{k_{21}! \cdots k_{2n_2}!}, \tag{11.3.4}$$

wobei sich die Summation über diejenigen $k_{21}, \ldots, k_{2n_2}$ erstreckt, welche den Relationen

$$\sum_j j k_{2j} = n_2, \quad \sum_j k_{2j} = k_2$$

genügen. Es gilt nun die Identität

$$(x + x^2 + \cdots)^{k_2} = x^{k_2} (1 + x + x^2 + \cdots)^{k_2} = \frac{x^{k_2}}{(1-x)^{k_2}} = x^{k_2} \sum_{m=0}^{\infty} \frac{(k_2 + m - 1)!}{(k_2 - 1)!\, m!} x^m.$$

Wir vergleichen die Koeffizienten von x^{n_2} auf der rechten und der linken Seite dieser Identität. Auf der rechten Seite ist dieser Koeffizient gleich

$$\binom{n_2 - 1}{k_2 - 1}, \tag{11.3.5}$$

während er auf der linken Seite gleich (11.3.4) ist. Aus den Formeln (11.3.3) bis (11.3.5) ergibt sich

$$\begin{aligned} P(k_{1j}, k_{2j}, n_1, n_2) &= \frac{k_1!}{k_{11}! \cdots k_{1n_1}!} G(k_1, k_2) p^{n_1} (1-p)^{n_2} \sum \frac{k_2!}{k_{21}! \cdots k_{2n_2}!} \\ &= \frac{k_1!}{k_{11}! \cdots k_{1n_1}!} \binom{n_2 - 1}{k_2 - 1} G(k_1, k_2) p^{n_1} (1-p)^{n_2}. \end{aligned} \tag{11.3.6}$$

Um die gesuchte Funktion $P(k_{1j}, n_1, n_2)$ zu erhalten, summieren wir (11.3.6) bezüglich k_2. Wenn wir die Beziehung zwischen K_1 und K_2 und die Definition der

Funktion $G(k_1, k_2)$ beachten, erhalten wir

$$(n_2 - 1)! \sum_{k_2=1}^{n_2} \frac{G(k_1, k_2)}{(k_2 - 1)!\,(n_2 - k_2)!}$$
$$= (n_2 - 1)! \left(\frac{2}{(k_1 - 1)!\,(n_2 - k_1)!} + \frac{1}{(k_1 - 2)!\,(n_2 - k_1 + 1)!} + \frac{1}{k_1!\,(n_2 - k_1 - 1)!} \right) = \binom{n_2 + 1}{k_1}.$$

Dies und die Formel (11.3.6) ergibt

$$P(k_{1j}, n_1, n_2)$$
$$= \frac{k_1!}{k_{11}! \cdots k_{1n_1}!}\, p^{n_1}(1 - p)^{n_2}(n_2 - 1)! \sum_{k_2=1}^{n_2} \frac{G(k_1, k_2)}{(k_2 - 1)!\,(n_2 - k_2)!}$$
$$= \frac{k_1!}{k_{11}! \cdots k_{1n_1}!} \binom{n_2 + 1}{k_1} p^{n_1}(1 - p)^{n_2}. \tag{11.3.7}$$

Einen ähnlichen Ausdruck erhält man für die Wahrscheinlichkeitsfunktion $P(k_{2j}, n_1, n_2)$.

C. Große Bedeutung für die Anwendungen besitzt die Wahrscheinlichkeitsfunktion $P(k_1, k_2, n_1, n_2)$ der Zufallsvariablen (K_1, K_2, N_1, N_2). Um diese Funktion zu finden, summieren wir (11.3.6) bezüglich k_{1j}, und zwar genauso, wie wir es in (11.3.3) taten, um (11.3.6) herzuleiten. Wenn wir also beachten, daß die Ausdrücke (11.3.4) und (11.3.5), in denen wir vorher k_2, n_2 und k_{2j} durch k_1, n_1 und k_{1j} ersetzt haben, einander gleich sind, so führt die Formel (11.3.6) zu

$$P(k_1, k_2, n_1, n_2) = \binom{n_1 - 1}{k_1 - 1}\binom{n_2 - 1}{k_2 - 1} G(k_1, k_2)\, p^{n_1}(1 - p)^{n_2}. \tag{11.3.8}$$

Aus der letzten Formel finden wir leicht durch Summation über k_2 bzw. k_1 die Wahrscheinlichkeitsfunktionen $P(k_1, n_1, n_2)$ und $P(k_2, n_1, n_2)$ der Zufallsvariablen (K_1, N_1, N_2) bzw. (K_2, N_1, N_2). Wir erhalten nämlich

$$P(k_1, n_1, n_2) = \binom{n_1 - 1}{k_1 - 1}\binom{n_2 + 1}{k_1} p^{n_1}(1 - p)^{n_2},$$
$$P(k_2, n_1, n_2) = \binom{n_1 + 1}{k_2}\binom{n_2 - 1}{k_2 - 1} p^{n_1}(1 - p)^{n_2}. \tag{11.3.9}$$

D. Schließlich wollen wir die Wahrscheinlichkeitsfunktion der Gesamtiterationsanzahl, d. h. der Zufallsvariablen $K = K_1 + K_2$ bestimmen. Hier unterscheiden wir zwei Fälle:

a) k ist eine gerade Zahl; dann muß $k_1 = k_2 = \frac{k}{2}$ sein. Aus Formel (11.3.8) ergibt sich, wenn wir $G\left(\frac{k}{2}, \frac{k}{2}\right) = 2$ berücksichtigen,

$$P(K = k) = P\left(\frac{k}{2}, \frac{k}{2}, n_1, n_2\right) = 2\binom{n_1 - 1}{\frac{k}{2} - 1}\binom{n_2 - 1}{\frac{k}{2} - 1} p^{n_1}(1 - p)^{n_2}; \tag{11.3.10}$$

b) k ist eine ungerade Zahl; dann ist entweder $k_1 = \frac{k+1}{2}$ und $k_2 = \frac{k-1}{2}$ oder $k_1 = \frac{k-1}{2}$ und $k_2 = \frac{k+1}{2}$. Da für diesen Fall $G(k_1, k_2) = 1$ ist, erhalten wir aus Formel (11.3.8)

$$P(K = k) = P\left(\frac{k+1}{2}, \frac{k-1}{2}, n_1, n_2\right) + P\left(\frac{k-1}{2}, \frac{k+1}{2}, n_1, n_2\right)$$

$$= \left[\binom{n_1 - 1}{\frac{k-1}{2}}\binom{n_2 - 1}{\frac{k-3}{2}} + \binom{n_1 - 1}{\frac{k-3}{2}}\binom{n_2 - 1}{\frac{k-1}{2}}\right] p^{n_1}(1 - p)^{n_2}. \tag{11.3.10'}$$

Aus (11.3.9) und (11.3.10) kann man die bedingten Wahrscheinlichkeitsfunktionen der dort auftretenden Zufallsvariablen K unter den Bedingungen $N_1 = n_1$ und $N_2 = n_2$ bestimmen. Wir haben nämlich

$$P(K_1 = k_1 \mid N_1 = n_1, N_2 = n_2) = \frac{\binom{n_1 - 1}{k_1 - 1}\binom{n_2 + 1}{k_1}}{\binom{n}{n_1}},$$

$$P(K_2 = k_2 \mid N_1 = n_1, N_2 = n_2) = \frac{\binom{n_1 + 1}{k_2}\binom{n_2 - 1}{k_2 - 1}}{\binom{n}{n_1}}. \tag{11.3.11}$$

Ähnlich haben wir für gerade bzw. ungerade k

$$P(K = k \mid N_1 = n_1, N_2 = n_2) = 2\,\frac{\binom{n_1 - 1}{\frac{k}{2} - 1}\binom{n_2 - 1}{\frac{k}{2} - 1}}{\binom{n}{n_1}},$$

$$P(K = k \mid N_1 = n_1,\, N_2 = n_2)$$
$$= \left[\frac{\binom{n_1-1}{\frac{k-1}{2}}\binom{n_2-1}{\frac{k-3}{2}}}{\binom{n}{n_1}} + \frac{\binom{n_1-1}{\frac{k-3}{2}}\binom{n_2-1}{\frac{k-1}{2}}}{\binom{n}{n_1}}\right]. \tag{11.3.12}$$

SWED und EISENHART [1] arbeiteten für n_1 und n_2, beide Werte kleiner als 21, Tafeln aus, die die Werte der Funktion $P(K < k \mid N_1 = n_1, N_2 = n_2)$ angeben.

Die oben angegebenen Ergebnisse aus dem Gebiet der Iterationstheorie stammen von BORTKIEWICZ [2], MOOD [1], STEVENS [1] u. a.

Beispiel 11.3.1. Es sei $P(X_j = a_1) = P(X_j = a_2) = 0{,}5$. In einer einfachen Stichprobe vom Umfang $n = 8$ befinden sich $n_1 = 5$ Glieder a_1 und $n_2 = 3$ Glieder a_2, die folgendermaßen angeordnet sind:

$$a_1, a_1, a_1, a_1, a_2, a_1, a_2, a_2.$$

Es gilt hier

$$k_{11} = k_{14} = 1, \quad k_{12} = k_{13} = k_{15} = 0, \quad k_{21} = k_{22} = 1, \quad k_{23} = 0,$$
$$k_1 = k_2 = 2, \quad k = 4, \quad n_1 = 5, \quad n_2 = 3, \quad n = 8.$$

Wie wir berechnen, ist $P(N_1 = 5, N_2 = 3) = 0{,}2234$. Weiter berechnen wir aus der Formel (11.3.3) die Wahrscheinlichkeit einer Gruppierung der Glieder a_1 und a_2, für die die Zufallsvariablen K_{1j}, K_{2j}, N_1 und N_2 die in diesem Beispiel beobachteten Werte annehmen. Diese Wahrscheinlichkeit beträgt 0,0319. Aus dem Kolmogoroffschen Gesetz der großen Zahlen folgt: Wenn wir mehrmals unabhängige einfache Stichproben zu 8 Elementen erheben, so erhalten wir in ungefähr 22 von 100 Fällen Stichproben mit 5 Gliedern a_1 und 3 Gliedern a_2. Dagegen enthalten nur ungefähr 3 von 100 Stichproben die von uns beobachteten Iterationsanzahlen.

Auf dieses Beispiel werden wir noch in 14.2 zurückkommen.

E. Ohne Beweis wollen wir noch zwei Ergebnisse angeben, die die Grenzverteilungen von Iterationen betreffen. WISHART und HIRSCHFELD [1] bewiesen, daß die Zufallsvariable K, deren Verteilung durch die Formeln (11.3.10) und (11.3.10') bestimmt ist, für $n \to \infty$ asymptotisch normal

$$N\left(2npq;\, 2\sqrt{npq(1-3pq)}\right)$$

verteilt ist. WALD und WOLFOWITZ [2] bewiesen, daß die durch (11.3.12) gegebene bedingte Verteilung für $n_1 = \alpha n_2$ $(\alpha > 0)$ und $n_1 \to \infty$ asymptotisch normal

$$N\left(\frac{2n_1}{1+\alpha};\; \sqrt{\frac{4\alpha n_1}{(1+\alpha)^3}}\right)$$

verteilt ist. Die letzte asymptotische Verteilung kann man schon für $n_1 > 20$ und $n_2 > 20$ benutzen.

Man kann auch Iterationen betrachten, die sich aus mehr als zwei Elementen zusammensetzen. Viele Ergebnisse auf diesem Gebiet findet der Leser in der Arbeit von MOOD [1].

Des weiteren kann man noch allgemeinere Iterationen betrachten, und zwar solche, bei denen die Zufallsvariablen X_j (siehe den Anfang von 11.2) unabhängig sind, jedoch eine homogene Markoffsche Kette bilden. Gewisse Ergebnisse auf diesem Gebiet findet man z. B. in der Arbeit von BABKIN, BELJAJEW und MAXIMOW [1].

11.4. Mittelwerte und Dispersionen der Iterationsanzahlen

An dieser Stelle wollen wir Mittelwerte und Dispersionen für einige der in 11.3 betrachteten Zufallsvariablen berechnen.

Aus (11.3.11) ergibt sich

$$E(K_1|N_1 = n_1, N_2 = n_2) = \sum_{k_1=1}^{n_1} k_1 \frac{\binom{n_1-1}{k_1-1}\binom{n_2+1}{k_1}}{\binom{n}{n_1}}$$

$$= (n_2+1) \sum_{k_1=1}^{n_1} \frac{\binom{n_1-1}{k_1-1}\binom{n_2}{k_1-1}}{\binom{n}{n_1}}. \tag{11.4.1}$$

Wir betrachten die Identität

$$(1+x)^{n_2}\left(1+\frac{1}{x}\right)^{n_1-1} \equiv \frac{(1+x)^{n_1+n_2-1}}{x^{n_1-1}}.$$

Entwickeln wir beide Seiten dieser Identität nach der binomischen Formel und betrachten wir die konstanten Glieder auf beiden Seiten, so haben wir auf der rechten Seite

$$\binom{n_1+n_2-1}{n_1-1},$$

während auf der linken Seite

$$\sum_{k_1=1}^{n_1} \binom{n_1-1}{k_1-1}\binom{n_2}{k_1-1}$$

steht.

Dies und (11.4.1) führen auf die Beziehung

$$E(K_1 | N_1 = n_1, N_2 = n_2) = \frac{1}{\binom{n}{n_1}} \frac{(n_2 + 1)(n - 1)!}{(n_1 - 1)!\, n_2!} = \frac{(n_2 + 1)\, n_1}{n}. \tag{11.4.2}$$

Vertauschen wir in dieser Formel die Rollen von n_1 und n_2 miteinander, so erhalten wir

$$E(K_2 | N_1 = n_1, N_2 = n_2) = \frac{(n_1 + 1)\, n_2}{n}. \tag{11.4.3}$$

Aus den beiden letzten Gleichungen folgt

$$\begin{aligned} &E(K | N_1 = n_1, N_2 = n_2) \\ &= E(K_1 | N_1 = n_1, N_2 = n_2) + E(K_2 | N_1 = n_1, N_2 = n_2) \\ &= \frac{2 n_1 n_2 + n}{n}. \end{aligned} \tag{11.4.4}$$

Um die bedingte Dispersion

$$D^2(K_1 | N_1 = n_1, N_2 = n_2)$$

zu finden, berechnen wir zuerst

$$E\big(K_1(K_1 - 1) | N_1 = n_1, N_2 = n_2\big).$$

Aus Formel (11.3.11) ergibt sich

$$E\big(K_1(K_1 - 1) | N_1 = n_1, N_2 = n_2\big) = n_2(n_2 + 1) \sum_{k_1=1}^{n_1} \frac{\binom{n_1 - 1}{k_1 - 1}\binom{n_2 - 1}{k_1 - 2}}{\binom{n}{n_1}}. \tag{11.4.5}$$

Wenn wir beide Seiten der Identität

$$(1 + x)^{n_2 - 1}\left(1 + \frac{1}{x}\right)^{n_1 - 1} \equiv \frac{(1 + x)^{n_1 + n_2 - 2}}{x^{n_1 - 1}}$$

nach der binomischen Formel entwickeln und die Koeffizienten von x^{-1} miteinander vergleichen, so erhalten wir

$$\sum_{k_1=1}^{n_1} \binom{n_1 - 1}{k_1 - 1}\binom{n_2 - 1}{k_1 - 2} = \binom{n_1 + n_2 - 2}{n_1 - 2}.$$

Diese Formel und (11.4.5) führen zu

$$E\big(K_1(K_1-1)\,|\,N_1=n_1,\ N_2=n_2\big)=\frac{n_2(n_2+1)\binom{n_1+n_2-2}{n_1-2}}{\binom{n}{n_1}}.$$

Aus der letzten Formel und aus (11.4.2) ergibt sich

$$D^2(K_1\,|\,N_1=n_1,\ N_2=n_2)=\frac{n_1(n_1-1)\,n_2(n_2+1)}{(n-1)n^2}. \tag{11.4.6}$$

Ähnlich erhalten wir

$$D^2(K_2|N_1=n_1,\ N_2=n_2)=\frac{n_1(n_1+1)n_2(n_2-1)}{(n-1)n^2}. \tag{11.4.7}$$

Die Formel

$$D^2(K\,|\,N_1=n_1,\ N_2=n_2)=\frac{2n_1n_2(2n_1n_2-n)}{(n-1)n^2} \tag{11.4.8}$$

geben wir ohne Beweis an, denn die zu ihr führenden elementaren Rechnungen sind etwas umständlich.

Aus (3.6.24), (11.4.2) und (11.4.6) ergibt sich

$$E(K_1)=E\left[\frac{n_1(n-n_1+1)}{n}\right]=np(1-p)+p^2,$$

$$D^2(K_1)=np(1-4p+6p^2-3p^3)+p^2(3-8p+5p^2).$$

Analog haben wir

$$E(K_2)=nq(1-q)+q^2,$$

$$D^2(K_2)=nq(1-4q+6q^2-3q^3)+q^2(3-8q+5q^2),$$

wobei $q=1-p$ ist.

Die Kovarianz μ_{11} der Zufallsvariablen K_1 und K_2 ist durch die Formel (siehe Mood [1])

$$\mu_{11}=-npq(3pq-1)-pq(2-5pq)$$

gegeben.

Weiter gilt

$$E(K)=E(K_1+K_2)=p^2+q^2+2npq,$$

$$D^2(K)=D^2(K_1)+D^2(K_2)+2\mu_{11}$$
$$=4npq(1-3pq)-2pq(3-10pq). \tag{11.4.9}$$

Insbesondere haben wir für $p = q = \frac{1}{2}$

$$E(K) = \frac{n+1}{2},$$

$$D^2(K) = \frac{n-1}{4}. \tag{11.4.10}$$

B. Zum Abschluß dieses Kapitels wollen wir den Leser darüber informieren, daß die Iterationstheorie bereits in den frühen Jahren der Entwicklung der modernen Wahrscheinlichkeitsrechnung die Aufmerksamkeit vieler Autoren auf sich gelenkt hat. Besonders wichtig auf diesem Gebiet ist das Buch von Bortkiewicz [2], und zwar sowohl wegen der dort enthaltenen Formeln als auch wegen der Diskussion mit Marbe [1] über dessen Versuch, die Anwendbarkeit der Wahrscheinlichkeitsrechnung durch Interpretation gewisser beobachteter Iterationen zu fördern. An dieser lebhaften Diskussion beteiligten sich Czuber [1], von Mises [2], Lexis [1] und Fryde [1].

11.5. Aufgaben und Ergänzungen

In den Aufgaben 1 und 2 sind die gleichen Überlegungen anzustellen wie bei der Herleitung der Formeln in 11.4.

1. Man zeige, daß die Gleichungen

$$E(K_{1j} \mid N_1 = n_1, N_2 = n_2) = \frac{(n_2+1)^{(2)}\, n_1^{(j)}}{n^{(j+1)}} \qquad (j = 1, 2, \ldots, n_1),$$

$$E(K_{2j} \mid N_1 = n_1, N_2 = n_2) = \frac{(n_1+1)^{(2)}\, n_2^{(j)}}{n^{(j+1)}} \qquad (j = 1, 2, \ldots, n_2)$$

gelten, wobei $a^{(l)} = a(a-1)\cdots(a-l+1)$ ist.

2. Man beweise für $m, j = 1, 2, \ldots, n_1$ und $s = 1, 2, \ldots, n_2$ die Gleichungen

$$D^2(K_{1j} \mid N_1 = n_1, N_2 = n_2) = \frac{n_2^{(2)}(n_2+1)^{(2)} n_1^{(2j)}}{n^{(2j+2)}} - \left[\frac{(n_2+1)^{(2)}\, n_1^{(j)}}{n^{(j+1)}}\right]^2,$$

$$\operatorname{cov}(K_{1j}, K_{1m}) = \frac{n_2^{(2)}(n_2+1)^{(2)}\, n_1^{(j+m)}}{n^{(j+m+2)}} - \frac{[(n_2+1)^{(2)}]^2\, n_1^{(j)}\, n_1^{(m)}}{n^{(j+1)}\, n^{(m+1)}}\Bigg],$$

$$\operatorname{cov}(K_{1j}, K_{1s}) = \frac{n_1^{(j+2)}\, n_2^{(s+2)}}{n^{(j+s+2)}} + 4\,\frac{n_1^{(j+1)}\, n_2^{(s+1)}}{n^{(j+s+1)}} + 2\,\frac{n_1^{(j)}\, n_2^{(s)}}{n^{(j+s)}} - \frac{(n_1+1)^{(2)}(n_2+1)^{(2)}\, n_1^{(j)}\, n_2^{(s)}}{n^{(j+1)}\, n^{(s+1)}}.$$

3. Man beweise: Ist $n_1 = \alpha n_2$ $(\alpha > 0)$ und strebt $n_1 \to \infty$, so ist die Grenzverteilung eines beliebigen Zufallsvektors $(K_{1j_1}, \ldots, K_{1j_l}, K_{2s_1}, \ldots, K_{2s_r})$ asymptotisch normal und hat die in den Aufgaben 1 und 2 angegebenen Mittelwerte, Varianzen und Kovarianzen (Mood [1]).
Hinweis. Man leite zunächst aus der Formel (11.3.3) die Wahrscheinlichkeitsfunktion $P(k_{1j}, k_{2j}, n_1, n_2 \mid N_1 = n_1, N_2 = n_2)$ her. Sodann betrachte man die standardisierten Zufallsvariablen $\left(K_{ij} - E(K_{ij})\right)/\sqrt{D^2(K_{ij})}$ und wende die Stirlingsche Formel an.

4. Wir bezeichnen mit R_{ij} die Anzahl der a_i-Iterationen der Länge $\geqq j$ $(i = 1, 2;$ $j = 1, 2, \ldots, n_i)$ und mit $R_j = R_{1j} + R_{2j}$ die allgemeine Anzahl der Iterationen der Länge $\geqq j$.

a) Man bestimme die Wahrscheinlichkeitsfunktionen dieser Zufallsvariablen.

b) Man beweise die Beziehungen

$$E(R_{1j} \mid N_1 = n_1, N_2 = n_2) = \frac{n_1^{(j)}(n_2 + 1)}{n^{(j)}} \qquad (j = 1, 2, \ldots, n_1),$$

$$E(R_{2j} \mid N_1 = n_1, N_2 = n_2) = \frac{n_2^{(j)}(n_1 + 1)}{n^{(j)}} \qquad (j = 1, 2, \ldots, n_2).$$

5. Man beweise, daß für $n_1 = n_2 = \frac{1}{2} n$ die folgenden Beziehungen gelten:

a) $E\left(K_{ij} \mid N_1 = N_2 = \frac{1}{2} n\right) \approx \frac{n}{2^{j+2}}$ $\qquad (i = 1, 2)$,

b) $E\left(R_{ij} \mid N_1 = N_2 = \frac{1}{2} n\right) \approx \frac{n}{2^{j+1}}$ $\qquad (i = 1, 2)$,

c) $E\left(R_j \mid N_1 = N_2 = \frac{1}{2} n\right) \approx \frac{n}{2^j}$.

12. SIGNIFIKANZTESTS

12.1. Der Begriff eines statistischen Tests

Jede Annahme über die unbekannte Verteilung einer Zufallsvariablen nennen wir eine *statistische Hypothese.* Wir wollen einige Beispiele für derartige Hypothesen bringen.

In Beispiel 12.1.1 legt die Hypothese nur den Wert eines unbekannten Verteilungsparameters einer Zufallsvariablen fest, von der wir wissen, daß sie normal verteilt ist.

Eine statistische Hypothese, die nur die Werte unbekannter Zahlenparameter einer Zufallsvariablen festlegt, nennt man eine *Parameterhypothese.*

Methoden, die zur Nachprüfung oder, wie man auch oft sagt, zur Verifikation der aufgestellten Hypothese H_0 dienen, nennt man *statistische Tests.* Einen Test, der zur Nachprüfung einer Parameterhypothese dient, nennt man einen *Parametertest.*

Stellen wir eine einzige Hypothese auf und dient der betrachtete Test nur dazu, nachzuprüfen, ob diese Hypothese nicht falsch ist, während gleichzeitig keine anderen Hypothesen untersucht werden, so spricht man von einem *Signifikanztest.* In diesem Kapitel wollen wir uns mit den Signifikanztests beschäftigen und an Beispielen zeigen, wie man statistische Hypothesen testet.

Beispiel 12.1.1. Das Merkmal X sei in einer gewissen Grundgesamtheit normal $N(m; 1)$ verteilt, der Mittelwert m sei unbekannt. Wir nehmen nun an, daß $m = 0$ gilt, mit anderen Worten, wir stellen eine statistische Hypothese auf, die wir mit H_0 bezeichnen und in der Form $H_0(m = 0)$ schreiben wollen.

Angenommen, die Hypothese $H_0(m = 0)$ sei richtig. Der betrachteten Gesamtheit entnehmen wir eine einfache Stichprobe vom Umfang 10, die den Mittelwert $\bar{x} = 1{,}01$ besitze. Dieser Wert stellt den beobachteten Wert $\bar{x}$ der Stichprobenfunktion $\bar{X}$ dar. Die Verteilung der Zufallsvariablen $\bar{X}$ in einfachen Stichproben aus einer normalen $N(0; 1)$ Gesamtheit ist uns genau bekannt. Nach Formel (9.3.5) ist in unserer Stichprobe die Variable $\bar{X}$ normal $N\left(0; \frac{1}{\sqrt{10}}\right)$ verteilt. Die Wahrscheinlichkeit dafür, einen Wert zu erhalten, der absolut genommen größer oder gleich dem beobachteten $\bar{x} = 1{,}01$ wäre, beträgt

$$P\left(|\bar{X}| \geqq 1{,}01\right) = P\left(|\bar{X}| \sqrt{10} \geqq 1{,}01 \sqrt{10}\right) \approx 2[1 - \Phi(3{,}05)] \approx 2(1 - 0{,}999)$$
$$= 0{,}002\,.$$

Ist also die Hypothese $H_0(m = 0)$ richtig, so kommt das beobachtete Ereignis in der Praxis äußerst selten vor; denn die Wahrscheinlichkeit dieses Ereignisses beträgt 0,002, d. h., in ungefähr 2 von 1000 Fällen würden wir einen Mittelwert $\bar{x}$ beobachten, dessen Absolutbetrag größer oder gleich 1,01 ist.

Danach ergibt sich folgende Frage: Können wir jetzt, da die Wahrscheinlichkeit des Ereignisses $|\overline{X}| \geqq 1{,}01$ sehr klein ist, als bewiesen annehmen, daß die Hypothese H_0 falsch ist? Wir müssen verneinen; denn ist auch die Wahrscheinlichkeit des Ereignisses $|\overline{X}| \geqq 1{,}01$ unter der Annahme, daß die Hypothese H_0 richtig ist, sehr klein, so kann doch ein derartiges Ereignis eintreten, wenn auch höchst selten.

Der Statistiker muß hier aber eine Entscheidung treffen, d. h., er muß sich für Annahme oder Ablehnung der Hypothese H_0 entscheiden. Dies tut er dadurch, daß er eine kritische Wahrscheinlichkeit α vorgibt und dann, wenn die Wahrscheinlichkeit des beobachteten Ereignisses (die Hypothese H_0 als richtig angenommen) kleiner oder gleich α ist, die Hypothese ablehnt.

Die so bestimmte kritische Wahrscheinlichkeit nennt man auch *Sicherheitswahrscheinlichkeit*. Die kritische Wahrscheinlichkeit α kann für verschiedene Probleme verschieden sein; sie ist im allgemeinen durch Konvention festgelegt. Gewöhnlich nimmt man $\alpha = 0{,}05$ oder $0{,}01$. Wählen wir aber $\alpha = 0{,}01$ (um so mehr also, wenn wir $\alpha = 0{,}05$ wählen), so müssen wir im Beispiel 12.1.1. die Hypothese $H_0(m = 0)$ ablehnen.

Beispiel 12.1.2. Einer normalen $N(m; 1)$ Gesamtheit, deren Mittelwert unbekannt ist, entnehmen wir eine einfache Stichprobe vom Umfang $n = 16$. Diese Stichprobe ergebe den Wert $\bar{x} = 0{,}1$. Wir wollen die Hypothese $H_0(m = 0)$ prüfen.

Ist die Hypothese H_0 richtig, so hat die Zufallsvariable $\overline{X}$ eine Normalverteilung $N\left(0; \frac{1}{4}\right)$. Wir berechnen die Wahrscheinlichkeit

$$P\left(|\overline{X}| \geqq 0{,}1\right) = P\left(4|\overline{X}| \geqq 0{,}4\right) = 2[1 - \Phi(0{,}4)] \approx 2 \cdot 0{,}345 = 0{,}690.$$

Die Wahrscheinlichkeit dafür, einen absolut genommen größeren oder gleichen Wert als den beobachteten zu erhalten, ist hier viel größer als $\alpha = 0{,}01$. Daraus folgt aber noch nicht, daß man die Hypothese H_0 ohne Vorbehalt als richtig betrachten kann.

Ein Signifikanztest erlaubt im allgemeinen nur, in einer Richtung zu entscheiden: Ist unter der Voraussetzung, daß die Hypothese H_0 richtig ist, die Wahrscheinlichkeit des beobachteten Ereignisses kleiner oder gleich α, so kann man die Hypothese H_0 ablehnen. Ist jedoch die Wahrscheinlichkeit des beobachteten Ereignisses größer als α, so kann man nur sagen, daß das Ergebnis des durchgeführten Versuchs nicht im Widerspruch zur Hypothese H_0 steht. Die Annahme einer Hypothese H_0 auf Grund eines einzigen Versuchs, dessen positiver Ausgang im Fall der Richtigkeit der Hypothese H_0 höchstens häufiger als einmal oder fünfmal in 100 Versuchen gewährleistet ist, ist weder durch die Wahrscheinlichkeitsrechnung begründet noch vernünftig. Wir heben diesen Umstand deshalb besonders hervor, weil manche Praktiker häufig den grundsätzlichen Fehler machen, die Hypothese als richtig zu betrachten, wenn der Signifikanztest die Hypothese nicht ablehnt.

Man kann jedoch statistische Testmethoden konstruieren, die es erlauben, Hypothesen zu bestätigen. Wir wollen sie in Kapitel 16 behandeln.

12.2. Parametertests für kleine Stichproben

Aus den Betrachtungen in 12.1 folgt, daß man das Wesen eines Parameters folgendermaßen charakterisieren kann:

Unbekannt sei der Verteilungsparameter Q eines Merkmals der Elemente einer Grundgesamtheit, während die Gestalt der Verteilungsfunktion bekannt sei. Wir stellen die Hypothese $H_0(Q = Q_0)$ auf. Der untersuchten Gesamtheit entnehmen wir eine n Elemente zählende Stichprobe und bestimmen aus ihr den beobachteten Wert u der Stichprobenfunktion U. Dann setzen wir eine kritische Wahrscheinlichkeit α fest, wobei α eine kleine Zahl ist, gewöhnlich 0,01 oder 0,05. Nun nehmen wir einen solchen Wert u_0 der Stichprobenfunktion U, daß unter Annahme der Gültigkeit von H_0 die Beziehung $P(|U| \geqq u_0) \leqq \alpha$ gilt. Erfüllt der beobachtete Wert u die Ungleichung $|u| \geqq u_0$, so lehnen wir die Hypothese H_0 ab. Wir stehen nämlich auf dem Standpunkt, daß der beobachtete Wert $|u|$ zu groß ist, als daß man sein Auftreten bei Gültigkeit von H_0 durch zufällige Schwankungen erklären könnte.

In einigen Problemen ist der Wert u_0 so zu wählen, daß $P(|U| \geqq u_0) \leqq \alpha$ oder $P(|U| < u_0) \leqq \alpha$ ist. Darüber, wie u_0 zu wählen ist, entscheidet der Charakter des betrachteten Problems. Eine ins Einzelne gehende Besprechung dieser Fragen findet der Leser in Kapitel 16.

Wie wir sehen, ist es äußerst wichtig, die Verteilung der Stichprobenfunktion U oder wenigstens den Wert u_0 zu kennen.

Bei der Durchführung eines Signifikanztests gehen wir verschieden vor, je nachdem, ob kleine oder große Stichproben vorliegen. Im ersten Fall benutzen wir die exakten Verteilungen der Stichprobenfunktionen und im zweiten die entsprechenden Grenzverteilungen.

Beispiel 12.2.1. Student [1] gab das folgende Beispiel an: Man untersucht die Wirksamkeit der Schlafmittel A und B an 10 Patienten, die an Schlaflosigkeit leiden. Die zusätzlichen Stunden an Schlaf, die ein Patient durch Einnehmen von A erlangt, seien mit X bezeichnet, die durch Einnehmen von B mit Y. Wir untersuchen die Zufallsvariable $Z = X - Y$. Die beobachteten Größen $x - y$ dieser Variablen stammen aus einer einfachen Stichprobe vom Umfang 10; die beobachteten Werte der Zufallsvariablen X, Y, Z wurden in Tab. 12.2.1 zusammengestellt. Man kann annehmen, daß die Zufallsvariable Z normal $N(m; \sigma)$ verteilt ist. Wir stellen die Hypothese $H_0(m = 0)$ auf, ohne zu präzisieren, wie groß der Wert von σ ist.

Der Tabelle 12.2.1 entnehmen wir $\bar{z} = 1{,}58$. Wir führen hier einen Signifikanztest für H_0 mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}01$ durch.

Hier ergibt sich die Frage, welche Stichprobenfunktion als Testgrundlage dienen soll. Auf den ersten Blick scheint es, daß man für Stichproben aus einer Gesamtheit die Stichprobenfunktion $\bar{Z}$ benutzen könnte, wie es im Beispiel 12.1.2 möglich war. Dies geht jedoch deshalb nicht, weil unter Voraussetzung der Richtigkeit von H_0 die Zufallsvariable $\bar{Z}$ zwar normal $N\left(0; \frac{\sigma}{\sqrt{10}}\right)$ verteilt, die Standardabweichung σ aber unbekannt ist. Die Stichprobe ist hier zu klein, um für die Standardabweichung σ den aus der Stichprobe berechneten

Tabelle 12.2.1.

Patient	x	y	$z = x - y$
1	1,9	0,7	1,2
2	0,8	−1,6	2,4
3	1,1	−0,2	1,3
4	0,1	−1,2	1,3
5	−0,1	−0,1	0,0
6	4,4	3,4	1,0
7	5,5	3,7	1,8
8	1,6	0,8	0,8
9	4,6	0,0	4,6
10	3,4	2,0	1,4

Wert $s = 1{,}167$ annehmen zu können. Die Zufallsvariable S konvergiert zwar stochastisch gegen σ (siehe 9.11), aber ihr für $n = 10$ beobachteter Wert kann bedeutend von σ abweichen. Deshalb wollen wir die Studentsche t-Verteilung benutzen.

Der in unserem Beispiel beobachtete Wert der Stichprobenfunktion t beträgt unter Voraussetzung der Richtigkeit der Hypothese $H_0(m = 0)$

$$t = \frac{\bar{z}}{s}\sqrt{n-1} = \frac{1{,}580}{1{,}167}\sqrt{9} = 4{,}06.$$

Aus den Tafeln der Studentschen t-Verteilung bestimmen wir denjenigen Wert t_0, für den die Beziehung $P(|t| \geqq t_0) = 0{,}01$ gilt. Für 9 Freiheitsgrade finden wir $t = 3{,}25$. Der beobachtete Wert t ist bedeutend größer als t_0, und deshalb lehnen wir die Hypothese H_0 ab. Das bedeutet, daß die Wirksamkeit der beiden Mittel nicht gleich ist. Das Schlafmittel A scheint wirksamer als das Mittel B zu sein.

Beispiel 12.2.2. In einer Fabrik produziert man Serien von Massenartikeln, deren Merkmal X normal $N(m; \sigma)$ verteilt sei; die Standardabweichung σ sei unbekannt. Unterliegt der Produktionsprozeß keinen wesentlichen Störungen, so bleibt die Standardabweichung σ für alle Serien dieselbe. Aus diesem Grunde stellt σ ein Maß für die Homogenität der Produktion oder für die Übereinstimmung des untersuchten Produktionsprozesses mit den technischen Normen dar.

Wir entnehmen einer Serie von Massenartikeln Stichproben und berechnen die Standardabweichung des x-Wertes in den Stichproben.

Wir wollen die Hypothese $H_0(\sigma = 4)$ mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}01$ testen. In einer einfachen Stichprobe vom Umfang 15, die dieser Gesamtheit entnommen wurde, erhielten wir nun den Wert $s = 4{,}5$. Wir fragen: Widerlegt dieser beobachtete Wert s unsere Hypothese? Muß man also annehmen, daß der beobachtete Wert s wesentlich größer als der mutmaßliche Wert σ ist, oder läßt sich diese Abweichung durch zufällige Schwankungen erklären, vorausgesetzt, die Hypothese gilt?

Hier benutzen wir die Verteilung der Stichprobenfunktion $Z = nS^2$, deren Dichte durch die Formel (9.5.12) bestimmt ist; diese Formel gibt die Dichte der χ^2-Verteilung mit $n - 1$ Freiheitsgraden an. Wenn wir $\sigma = 4$ annehmen, erhalten wir den Wert

$$\frac{z}{\sigma^2} = \frac{15 \cdot 4{,}5^2}{4^2} \approx 19.$$

Aus den Tafeln der χ^2-Verteilung ersehen wir, daß für 14 Freiheitsgrade $\alpha = 0{,}01$ die Wahrscheinlichkeit der Ungleichung $\frac{z}{\sigma^2} \geqq 29{,}1$ darstellt. Der von uns beobachtete Wert 19 widerlegt bei der Wahrscheinlichkeit $\alpha = 0{,}01$ die Hypothese H_0 nicht.

Beispiel 12.2.3. In einer Warenladung befinden sich fehlerlose und fehlerhafte Stücke, wobei der Ausschußkoeffizient p unbekannt sei. Wir stellen die Hypothese $H_0(p = 0{,}10)$ auf. Diese Hypothese besagt, daß die Ladung zu 10% aus Ausschuß besteht.

Wir entnehmen dieser Ladung eine einfache Stichprobe vom Umfang $n = 30$. Dem Auftreten eines fehlerhaften Stückes ordnen wir die Zahl 1, dem Auftreten eines fehlerlosen Stückes die Zahl 0 zu. Man fand nun in der Stichprobe 4 Ausschußstücke. Widerlegt diese Probenuntersuchung die Hypothese H_0 bei Vorgabe einer Sicherheitswahrscheinlichkeit von $\alpha = 0{,}05$?

Wir wollen die binomial verteilte Stichprobenfunktion

$$X = \sum_{k=1}^{30} X_k$$

untersuchen, wobei X_k Zufallsvariable mit derselben Null-Eins-Verteilung sind. Ist die Hypothese H_0 richtig, so ist die Wahrscheinlichkeitsfunktion dieser Zufallsvariablen durch

$$P(X = r) = \binom{30}{r} 0{,}1^r \cdot 0{,}9^{30-r} \qquad (r = 0, 1, \ldots, 30)$$

bestimmt. Daraus folgt

$$P(X \geqq 4) = \sum_{r=4}^{30} \binom{30}{r} 0{,}1^r \cdot 0{,}9^{30-r}. \tag{12.2.1}$$

Um diese Summe zu bestimmen, berechnen wir nur

$$P(X = 0) = 0{,}9^{30},$$

$$P(X = 1) = \frac{30!}{29!} 0{,}1 \cdot 0{,}9^{29} = 30 \cdot 0{,}1 \cdot 0{,}9^{29},$$

$$P(X = 2) = \frac{30!}{2!\,28!} 0{,}1^2 \cdot 0{,}9^{28} = 15 \cdot 29 \cdot 0{,}1^2 \cdot 0{,}9^{28},$$

$$P(X = 3) = \frac{30!}{3!\,27!} 0{,}1^3 \cdot 0{,}9^{27} = 5 \cdot 29 \cdot 28 \cdot 0{,}1^3 \cdot 0{,}9^{27}.$$

Die Summe dieser vier Wahrscheinlichkeiten ist

$$0{,}9^{27}[0{,}9^3 + 3 \cdot 0{,}9^2 + 4{,}35 \cdot 0{,}9 + 4{,}06] = 0{,}9^{27} \cdot 11{,}134 \approx 0{,}6473.$$

Somit ist die gesuchte Wahrscheinlichkeit gleich $1 - 0{,}6473 = 0{,}3527$. Daraus ersehen wir, daß der hier angewandte Test die Hypothese H_0 nicht widerlegt.

Beispiel 12.2.4. Die dem Buch von Bojew [1] entnommene Tabelle 12.2.2 enthält in ihrer Kopfzeile die Baumwollnummer X, die umgekehrt proportional dem in Zoll gemessenen Fadendurchmesser ist. In der ersten Spalte ist die Widerstandskraft Y des Fadens, gemessen

in Gramm, angegeben: Um die Widerstandskraft der Fäden zu prüfen, hat man 60 unabhängige Probeuntersuchungen für verschiedene Fadennummern angestellt. Die Kästchen der Tafel enthalten die Anzahl der beobachteten Wertepaare (x, y).

Wir nehmen an, die Variable (X, Y) habe eine zweidimensionale Normalverteilung, und stellen die Hypothese $H_0(\varrho = 0)$ auf; dabei bedeutet ϱ den Korrelationskoeffizienten der Variablen X und Y.

Wir wollen diese Hypothese an Hand der Ergebnisse, die wir aus der Probeuntersuchung erhielten, testen. So, wie die Zahlen in Tabelle 12.2.2 gruppiert sind, vermuten wir, daß der Test die Hypothese widerlegen wird.

Tabelle 12.2.2.

y \ x	4100	4300	4500	4700	4900	5100	5300	5500	zusammen
6,75	1	1							1
6,25		2	2	1					6
5,75		1	3	4	2	3			13
5,25		3	5	7	1	1			17
4,75			2	5	5	3	2		17
4,25					1	2	2		5
3,75								1	1
zusammen	1	7	12	17	9	9	4	1	60

Wir berechnen den beobachteten Wert r des Korrelationskoeffizienten R. Mit $\bar{x}$ und $\bar{y}$ bezeichnen wir die beobachteten Werte von $\bar{X}$ bzw. $\bar{Y}$. Es ergibt sich $\bar{x} = 4746{,}67$, $\bar{y} = 5{,}23$.

Nun benutzen wir Tabelle 12.2.3, deren Kopf die Werte $x - \bar{x}$ enthält, während in der ersten Spalte die Werte $y - \bar{y}$ eingetragen sind. Die einzelnen Kästen dieser Tabelle enthalten das Produkt $(x - \bar{x})(y - \bar{y})$, multipliziert mit der Beobachtungszahl aus Tabelle 12.2.2. Wir erhalten

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} = -0{,}61.$$

Um die Hypothese H_0 nachzuprüfen, benutzen wir die Tatsache, daß für einfache Stichproben vom Umfang n aus einer zweidimensionalen Gesamtheit mit $\varrho = 0$ die Stichprobenfunktion

$$\frac{R}{\sqrt{1 - R^2}} \sqrt{n - 2}$$

eine Studentsche t-Verteilung mit $n - 2$ Freiheitsgraden hat (vgl. 9.9). Der von uns beobachtete Wert dieser Stichprobenfunktion beträgt $-5{,}86$. Die Studentsche Stichprobenfunktion t ist für 58 Freiheitsgrade angenähert normal $N(0; 1)$ verteilt, die Wahrscheinlichkeit $P(|t| \geqq 5{,}86)$ ist also äußerst klein. Wäre also die Hypothese $H_0(\varrho = 0)$ richtig, so hätten wir ein äußerst seltenes Ereignis beobachtet, dessen Wahrscheinlichkeit viel kleiner als die von uns angenommene Sicherheitswahrscheinlichkeit α wäre. Daher lehnen wir die Hypothese H_0 ab.

Tabelle 12.2.3.

$y-\bar{y}$ \ $x-\bar{x}$	−646,67	−446,67	−246,67	−46,67	153,33	353,33	553,33	753,33
1,52		−678,94						
1,02	−659,60	−911,21	−503,21	−47,60				
0,52		−232,27	−348,81	−97,07	159,46	551,19		
0,02		−26,80	−24,67	−6,53	3,07	7,07		
−0,48			236,80	112,01	−367,99	−508,80	−531,20	
−0,98					−150,26	−692,53	1084,53	
−1,48								−1114,93

Beispiel 12.2.5. Die Tabelle 12.2.4 enthält die Kartoffel- und Eierpreise für den freien Handel, die im September 1947 in 14 polnischen Städten beobachtet wurden. Mit X bzw. Y seien die Kartoffel- bzw. Eierpreise bezeichnet. Aus den Angaben dieser Tabelle berechnen wir die Gleichung der Regressionsgeraden von Y nach X.

Tabelle 12.2.4[1])

Städte	Freihandelspreis in Złoty im September 1947 für	
	1 kg Kartoffeln	1 Ei
Warschau	11,4	12,8
Łódź	12,3	15,0
Kielce	15,0	12,5
Lublin	11,5	13,5
Białystok	5,0	11,0
Olsztyn	7,0	12,5
Gdańsk	11,0	14,0
Bydgoszcz	10,0	12,0
Szczecin	10,5	13,5
Poznań	10,5	14,3
Wrocław	12,3	13,5
Katowice	11,8	13,5
Kraków	17,0	13,3
Rzeszów	12,0	12,0

[1]) Quelle: Wiadomości Statystyczne.

Aus Tabelle 12.2.4 erhalten wir

$$\bar{x} = 11{,}24, \quad \bar{y} = 13{,}10, \quad s_1^2 = 7{,}79, \quad s_2^2 = 1{,}01, \quad r = 0{,}4.$$

Die Formel (9.10.1) ergibt

$$a = r\frac{s_2}{s_1} = 0{,}144, \quad b = \bar{y} - a\bar{x} = 11{,}48.$$

Die gesuchte Regressionsgerade hat also die Form

$$Y = 0{,}144\,X + 11{,}48.$$

Wenn wir annehmen, daß (X, Y) normal verteilt ist, können wir einen auf die Formel (9.10.15) gestützten Signifikanztest anwenden, um festzustellen, ob die Hypothese H_0 gilt, daß der wirkliche Wert des Regressionskoeffizienten α gleich 0 ist. Wir haben

$$t_1 = \frac{s_1}{s_2}\,\frac{\sqrt{n-2}}{\sqrt{1-r^2}}\,(a-\alpha) = \frac{2{,}791}{1{,}005}\,\frac{\sqrt{12}}{\sqrt{0{,}84}}\;0{,}144 \approx 1{,}61.$$

Da die Anzahl der Freiheitsgrade hier 12 beträgt, finden wir aus den Tafeln der Studentschen t-Verteilung, daß die Wahrscheinlichkeit dafür, einen nicht kleineren absoluten t-Wert als den beobachteten zu erhalten, größer als 0,05 ist. Es besteht also keine Notwendigkeit, die Hypothese H_0 aufzugeben. Dieses Ergebnis stimmt gut mit der Erfahrung überein, da Eier kein Ersatz für Kartoffeln sind.

Beispiel 12.2.6. Wir wenden uns wieder dem Beispiel 12.2.1 zu. An jedem Patienten wurde die Wirksamkeit zweier Schlafmittel ausprobiert, und wir betrachteten die Differenz $x - y$ als eine einzige Beobachtung an einem Patienten. Nun stellen wir uns vor, daß der Versuch anders durchgeführt wurde, und zwar, daß die Schlafmittel A und B jedes für sich an zwei aus je 10 Personen bestehenden Gruppen ausprobiert werden. Wir hätten dann zwei voneinander unabhängige einfache Stichproben vor uns. Die eine besteht aus 10 Beobachtungen der Zufallsvariablen X, die andere aus 10 Beobachtungen der Zufallsvariablen Y. Wir nehmen an, X und Y seien normal $N(m_1; \sigma)$ bzw. $N(m_2; \sigma)$ verteilt, und stellen die Hypothese $H_0(m_1 = m_2)$ auf, d. h., wir behaupten, die Wirksamkeit beider Mittel sei gleich. Um diese Hypothese zu verifizieren, benutzen wir die durch Formel (9.6.8) bestimmte Verteilung der Stichprobenfunktion U.

Ist die Hypothese $H_0(m_1 = m_2)$ richtig, so haben wir es mit Stichproben aus derselben normalen Gesamtheit zu tun. Um aus dem Beobachtungsmaterial den Wert u zu berechnen, entnehmen wir zuerst der Tabelle 12.2.1 die Werte

$$\bar{x} = \frac{1}{10}\sum x = 2{,}33, \qquad \bar{y} = \frac{1}{10}\sum y = 0{,}75,$$

$$10 s_1^2 = \sum (x-\bar{x})^2 = 36{,}1, \qquad 10 s_2^2 = \sum (y-\bar{y})^2 = 28{,}9$$

und erhalten

$$u = \frac{\bar{x}-\bar{y}}{\sqrt{10(s_1^2+s_2^2)}}\sqrt{\frac{100}{20}\cdot 18} \approx 1{,}86.$$

Wie man weiß (siehe 9.6), hat die Variable U eine Studentsche t-Verteilung mit 18 Freiheitsgraden. Den Tafeln der t-Verteilung entnehmen wir, daß die Wahrscheinlichkeit dafür, einen absoluten t-Wert nicht kleiner als 1,86 zu erhalten, gleich $P = 0{,}08$ ist. Das erhaltene Ergebnis widerspricht also nicht der Hypothese $H_0(m_1 = m_2)$, sogar nicht einmal bei der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$.

Beim Vergleich dieses Schlusses mit dem im Beispiel 12.2.1 formulierten Schluß mag es dem Leser unverständlich erscheinen, wieso beide Schlüsse einander entgegengesetzt sind. Diese Tatsache kann man aber einfach und intuitiv

erklären. Im Beispiel 12.2.1 hat man die Wirksamkeit beider Mittel an denselben Patienten untersucht. Ist also die durchschnittliche Wirksamkeit beider Mittel dieselbe, dann müßte man nur kleine Differenzen in den Auswirkungen beobachten. Die beobachteten Differenzen $\bar{x} - \bar{y}$ kann man also nicht durch zufällige Schwankungen (bei der angenommenen Sicherheitswahrscheinlichkeit) erklären. Im Beispiel 12.2.6 untersucht man dagegen den Unterschied in der Wirksamkeit dieser beiden Mittel an zwei verschiedenen Patientengruppen. Hier kann man auch dann, wenn die Hypothese $H_0(m_1 = m_2)$ richtig ist, größere Differenzen erwarten.

Im Beispiel 12.2.6 betrachteten wir zwei normale Grundgesamtheiten mit den unbekannten Mittelwerten m_1 und m_2 und den unbekannten Standardabweichungen σ_1 und σ_2. Wir hatten als sicher, d. h. als keines Beweises bedürftig angenommen, daß $\sigma_1 = \sigma_2 = \sigma$ ist, wobei der Wert von σ nicht näher spezifiziert wurde, und wir formulieren die Hypothese $H_0(m_1 = m_2)$. Somit haben, wenn die Hypothese H_0 richtig ist, die in diesem Beispiel betrachteten Zufallsvariablen X und Y die gleiche Normalverteilung. Die Hypothese H_0 wurde auf der Basis der Studentschen t-Verteilung überprüft. In der Praxis ist es jedoch häufig notwendig, entweder die Hypothese $H_1(m_1 = m_2, \sigma_1 = \sigma_2)$ oder die Hypothese $H_2(m_1 = m_2)$ ohne die Annahme $\sigma_1 = \sigma_2$ zu überprüfen. Betrachten wir somit die Hypothese H_1, so muß die Annahme $\sigma_1 = \sigma_2$ überprüft werden, betrachten wir hingegen die Hypothese H_2, so schließen wir die Möglichkeit $\sigma_1 \neq \sigma_2$ nicht aus, d. h., daß X und Y verschiedene Normalverteilungen haben.

Das Problem der Überprüfung der Hypothesen H_1 und H_2 war Gegenstand vieler Diskussionen und Kontroversen. Wir erwähnen hier die Arbeiten von Fisher [7], [12], Neyman und Pearson [2], Sukhatme [2], Hsu [1] und Welch [1].

12.3 Parametertests für große Stichproben

Wir wollen an Beispielen zeigen, wie man bei Signifikanztests die Grenzverteilung von Stichprobenfunktionen benutzen kann.

Beispiel 12.3.1. Die Verteilung des Merkmals X einer Gesamtheit sei unbekannt; es möge aber die Standardabweichung $\sigma = 5$ dieser Verteilung gegeben sein. Dieser Gesamtheit entnehmen wir eine einfache Stichprobe vom Umfang 50. Der Mittelwert m des Merkmals X der Gesamtheit sei ebenfalls unbekannt. Wir stellen die Hypothese $H_0(m = 0)$ auf. Die Stichprobe ergab den Wert $\bar{x} = 2$. Die Hypothese H_0 ist zu testen.

Aus der Existenz einer endlichen Standardabweichung folgt (siehe 9.11): Ist die Hypothese H_0 richtig, so ist die Stichprobenfunktion $\bar{X}$ asymptotisch normal $N\left(0; \frac{1}{\sqrt{2}}\right)$ verteilt. Aus den Tafeln der Normalverteilung finden wir

$$P(|\bar{X}| \geqq 2) = P(|\bar{X}|\sqrt{2} > 2\sqrt{2}) \approx P(|\bar{X}|\sqrt{2} \geqq 2{,}82) < 0{,}01.$$

Wir lehnen also die Hypothese H_0 ab.

In Beispiel 12.3.1 setzen wir voraus, daß die Standardabweichung der Grundgesamtheit bekannt ist. Gewöhnlich ist sie aber in praktischen Fällen nicht

bekannt. Ist die Stichprobe jedoch umfangreich genug, so können wir die Standardabweichung aus der Stichprobe bestimmen, also $\sigma = s$ annehmen. Dies ist aber nur dann möglich, wenn die Stichprobe wenigstens 100 Elemente enthält.

Beispiel 12.3.2. Wir haben eine einfache Stichprobe vom Umfang $n = 150$ Elemente. Aus ihr bestimmen wir die Werte $\bar{x} = 0{,}4$ und $s = 4$. Wir kennen weder den Mittelwert noch die Standardabweichung der Gesamtheit. Ist die Hypothese $H_0(m = 0)$ abzulehnen?

Wenn wir $\sigma = s$ nehmen und die Hypothese H_0 richtig ist, dann ist die Zufallsvariable $\bar{X}$ asymptotisch normal $N\left(0; \frac{4}{\sqrt{150}}\right)$ verteilt. Wir erhalten

$$P\left(|\bar{X}| \geqq 0{,}4\right) = P\left(\frac{|\bar{X}|}{0{,}33} \geqq 1{,}2\right) = 0{,}23.$$

Es besteht also kein Grund, die Hypothese H_0 abzulehnen.

Wie wir sehen, benutzt man in umfangreichen Stichproben Grenzverteilungen und nimmt außerdem in einigen Fällen noch an, daß die Werte der anderen Parameter (außer dem Parameter, der nachgeprüft werden soll) den aus der Stichprobe errechneten Parameterwerten gleich sind. Dazu bemerken wir, daß die Aussage, ob eine Stichprobe als groß oder klein anzusehen ist, davon abhängt, welchen Parameter die Hypothese betrifft, und davon, ob die Stichprobenfunktion, die die Grundlage des Parametertestes bildet, schnell oder langsam stochastisch gegen den gesuchten Gesamtheitsparameter konvergiert. Bestimmt man aus der Stichprobe auch noch andere Parameter (wie dies im Beispiel 12.3.2 der Fall ist), so ist für sie derselbe Umstand zu beachten.

Beispiel 12.3.3. Ein Merkmal, das in der allgemeinen Landwirtschaftszählung untersucht wurde, ist der Besitz eines Arbeitspferdes. Wir wollen feststellen, wieviel landwirtschaftliche Betriebe in den Woiwodschaften A und B ohne Arbeitspferd sind. Dazu wählen wir nach dem Bernoullischen Schema $n_1 = 1500$ Zählformulare der Woiwodschaft A und $n_2 = 1800$ der Woiwodschaft B. Der Anteil der Landwirtschaften ohne Pferd beträgt in den Stichproben, die den Woiwodschaften A bzw. B entnommen sind,

$$p_1 = \frac{300}{1500} = 0{,}200 \quad \text{bzw.} \quad p_2 = \frac{320}{1800} = 0{,}178$$

Wir stellen die Hypothese H_0 auf, daß in beiden Woiwodschaften der Anteil der Landwirtschaften ohne Pferd der gleiche ist und p beträgt und daß nach dieser Hypothese die in diesen Stichproben beobachteten Unterschiede zwischen p_1 und p_2 zufällige Schwankungen sind.

Diese Hypothese wollen wir mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ testen.

Wir ordnen dem darin bestehenden Ereignis, daß wir ein Formular einer Landwirtschaft ohne Pferd auslosen, die Zahl 1 und dem entgegengesetzten Ereignis die Zahl 0 zu. Also haben wir es hier mit Zufallsvariablen X_k, die eine Null-Eins-Verteilung haben, zu tun. Der Wert $p_1 = 0{,}200$ stellt den beobachteten Wert der Zufallsvariablen Y in einem Bernoullischen Schema dar; dabei ist $Y = \frac{1}{n_1} \sum X_k$, $n_1 = 1500$, während $p_2 = 0{,}178$ der beobach-

tete Wert der Zufallsvariablen Z in einem Bernoullischen Schema ist: $Z = \frac{1}{n_2} \sum X_k$, $n_2 = 1800$.

Ist die Hypothese H_0 richtig, so ist $E(Z) = E(Y) = p$. Die Standardabweichungen der Zufallsvariablen Y bzw. Z sind auf Grund der Formeln (5.2.8) gleich

$$\sigma_y = \sqrt{\frac{p(1-p)}{1500}} \quad \text{bzw.} \quad \sigma_z = \sqrt{\frac{p(1-p)}{1800}}.$$

Die Zufallsvariablen Y und Z sind voneinander unabhängig, also hat die Zufallsvariable $Y - Z$ auf Grund der Formeln (3.2.14) und (3.6.17) die Standardabweichung

$$\sigma = \sqrt{\sigma_y^2 + \sigma_z^2} = \sqrt{\frac{11p(1-p)}{9000}}.$$

Die Zufallsvariable $Y - Z$ hat eine asymptotische Normalverteilung

$$N\left(0; \sqrt{\frac{11p(1-p)}{9000}}\right).$$

Würden wir den Wert von p kennen, so könnten wir den genauen Wert von σ berechnen. Da die Stichprobenumfänge aber groß sind, können wir für den Wert p das folgende Mittel von p_1 und p_2 nehmen, nämlich

$$p = \frac{1500\,p_1 + 1800\,p_2}{3300} = 0{,}188.$$

Wir erhalten $\sigma = 0{,}0137$. Der von uns beobachtete Wert der Zufallsvariablen $Y - Z$ ist $p_1 - p_2 = 0{,}022$. Aus den Tafeln der Normalverteilung lesen wir ab:

$$P(|Y - Z| \geqq 0{,}022) = P\left(\frac{|Y - Z|}{0{,}0137} \geqq \frac{0{,}022}{0{,}0137}\right) = P\left(\frac{|Y - Z|}{0{,}0137} \geqq 1{,}61\right) = 0{,}107.$$

Wie wir sehen, führt hier der Signifikanztest mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ nicht zur Ablehnung der Hypothese H_0. Die Ergebnisse der ausgeführten Stichproben schließen nicht die Möglichkeit aus, daß der Prozentsatz der Landwirtschaften ohne Pferd für beide Woiwodschaften der gleiche ist.

Beispiel 12.3.4. Man untersuchte den Roggenbrotverbrauch eines Normalverbrauchers[1]) für Arbeiterfamilien in Warschau und Łódź und in Górny Śląsk (Oberschlesien). So hat man in Warschau und Łódź $n_1 = 65$ Familien und im oberschlesischen Industriegebiet $n_2 = 57$ Familien untersucht. Wir bezeichnen den Brotverbrauch in Kilogramm pro Normalverbraucher mit X. Die untersuchten Familiengruppen kann man als einfache Stichprobe aus den Gesamtheiten der Arbeiterfamilien in Warschau und Łódź bzw. der oberschlesischen Familien betrachten. Unter m_1 bzw. m_2 verstehen wir den mittleren Roggenbrotverbrauch in Warschau und Łódź bzw. in Oberschlesien.

[1]) Unter einem *Normalverbraucher* versteht man einen Mann, der 18 Jahre oder älter ist; der Nahrungsmittelverbrauch von Personen, die einer anderen Alters- oder Geschlechtsgruppe angehören, wird nach speziellen Skalen umgerechnet.

Die beobachteten Stichprobenmittelwerte betragen $\bar{x}_1 = 9{,}95$, $\bar{x}_2 = 10{,}28$. Wir stellen die Hypothese auf, daß die durchschnittlichen Verbrauchswerte m_1 und m_2 einander gleich sind.

Den Wert $\bar{x}_1 - \bar{x}_2$ können wir als beobachteten Wert der Zufallsvariable $\bar{X}_1 - \bar{X}_2$ ansehen, deren Verteilung wegen der Größe des Stichprobenumfanges als normal angenommen werden kann. Ist die Hypothese H_0 richtig, so haben wir es hier mit der Normalverteilung $N\left(0;\ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ zu tun, wobei σ_1 bzw. σ_2 die Standardabweichungen der Zufallsvariablen X in Warschau und Łódź bzw. in Oberschlesien sind. Wir kennen diese Werte nicht. Da aber die Stichprobenumfänge *nicht klein* sind, nehmen wir für sie die aus den Stichproben erhaltenen Abweichungen. Diese betragen $s_1 = 4{,}22$ und $s_2 = 6{,}01$, die Standardabweichung der Variablen $\bar{X}_1 - \bar{X}_2$ beträgt also 0,95. Da in unserem Beispiel $\bar{x}_1 - \bar{x}_2 = -0{,}33$ ist, finden wir aus den Tafeln der Normalverteilung die Wahrscheinlichkeit dafür, daß eine normale Variable mit der Verteilung $N(0; 0{,}95)$ absolut genommen nicht kleiner als 0,33 ist, den Wert 0,7264. Testen wir die Hypothese H_0 mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$, so besteht kein Grund zu ihrer Ablehnung.

Beispiel 12.3.5. Sittig und Freudenthal [1] beschreiben die Ergebnisse einer in Holland an 5001 Frauen durchgeführten anthropometrischen Untersuchung. Diese Untersuchungen dienten den Bedürfnissen der Konfektionsindustrie, sie bezweckten die Auswahl einer Reihe von Mannequins, so daß die nach diesen Modellen hergestellten Bekleidungsstücke und Schuhe den holländischen Frauen möglichst gut paßten. Unter den untersuchten Merkmalen befanden sich unter anderen die Körpergröße Y und die Länge des Mittelfingers X. Aus den Messungen fand man für den Korrelationskoeffizienten zwischen diesen Merkmalen den Wert $r = 0{,}5062$.

Wir stellen die Hypothese $H_0(\varrho = 0)$ auf, d. h., wir nehmen an, daß der Korrelationskoeffizient zwischen den betrachteten Merkmalen der Gesamtheit gleich 0 ist. Da die Stichprobe sehr groß ist, können wir annehmen, daß R normal verteilt ist, und die Formeln (9.9.9) für $E(R)$ und $D^2(R)$ benutzen; dabei ersetzen wir in ihnen den unbekannten Wert ϱ durch 0. Ist die Hypothese H_0 richtig, so hat R eine in Annäherung normale $N(0; 0{,}0141)$ Verteilung. Wir können leicht feststellen, daß die Verifizierung der Hypothese $H_0(\varrho = 0)$ sogar mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}01$ zur Ablehnung dieser Hypothese führt; der Korrelationskoeffizient ϱ ist also wesentlich von Null verschieden.

12.4. Anpassungstests. Der χ^2-Test

A. Es sei $F(x)$ die unbekannte Verteilungsfunktion der Zufallsvariablen X.

Eine statistische Hypothese, die die unbekannte Gestalt der Verteilungsfunktion $F(x)$ festlegt, nennt man *nichtparametrisch*.

Ein Verfahren, das zur Verifizierung einer nichtparametrischen Hypothese dient, nennt man einen *Anpassungstest* oder *nichtparametrischen Test*.

Um eine nichtparametrische Hypothese nachzuprüfen, gehen wir folgendermaßen vor. Wir stellen die Hypothese H_0 über die Verteilungsfunktion $F(x)$ auf. Der Grundgesamtheit entnehmen wir eine Stichprobe vom Umfang n und untersuchen in ihr die Verteilung des Merkmals X. Diese empirische Verteilung

wird im allgemeinen von der durch die Hypothese H_0 festgelegten Gesamtverteilung verschieden sein, und auch dann, wenn die Hypothese richtig ist, ist im allgemeinen in verschiedenen Stichproben aus derselben Gesamtheit die empirische Verteilung verschiedenartig. Ist nun die Hypothese richtig, so können sich diese empirischen Verteilungen nicht wesentlich von der hypothetischen Verteilung unterscheiden. Um diese Erscheinung quantitativ zu erfassen, müssen wir eine Stichprobenfunktion U konstruieren, die ein Maß für die Abweichung (oder Divergenz) der empirischen Verteilung von der hypothetischen darstellt. Die weitere Verfahrensweise ist dann analog zu derjenigen, die bei Parametertests angewandt wurde: Wir bestimmen einen Wert u_0 derart, daß dann, wenn die Hypothese H_0 richtig ist, $P(U \geqq u_0) \leqq \alpha$ gilt, wobei α die Sicherheitswahrscheinlichkeit ist. Ist der aus der untersuchten Stichprobe berechnete Wert u größer oder gleich u_0, so lehnen wir die Hypothese H_0 ab. Wenn dagegen $u < u_0$ ist, so besteht kein Grund zur Ablehnung von H_0, und wir müssen über die Annahme dieser Hypothese durch weitere Untersuchungen entscheiden.

B. Oft wird die von K. PEARSON [5] stammende Stichprobenfunktion χ^2 als Maß für die Abweichung benutzt. Wir erinnern daran, daß eine Stichprobenfunktion desselben Namens schon in Kapitel 9 besprochen wurde. Wie wir sehen werden, sind die Verteilungen der gleichbenannten Stichprobenfunktion eng miteinander verknüpft. Man darf aber nicht vergessen, daß ihre Definitionen völlig verschieden sind.

Wir wollen jetzt die Pearsonsche Stichprobenfunktion χ^2 definieren. Es sei $F(x)$ die gegebene Verteilungsfunktion der Grundgesamtheit. Wir zerlegen die x-Achse in r paarweise punktfremde Mengen S_k. Es sei

$$\pi_k \ (k = 1, 2, \ldots, r)$$

die Wahrscheinlichkeit dafür, daß die Zufallsvariable X einen Wert aus S_k annimmt. Wenn wir für S_k die Intervalle $[a_k, a_{k+1})$ nehmen, erhalten wir

$$\pi_k = F(a_{k+1}) - F(a_k). \tag{12.4.1}$$

Die Zahlen π_k wollen wir *theoretische Häufigkeiten* nennen.

Es mögen nun n voneinander unabhängige Beobachtungen $x_1, x_2, \ldots, x_n$ der Zufallsvariablen X vorliegen. Diese Beobachtungen teilen wir in r Gruppen ein. Zur k-ten Gruppe zählen wir diejenigen Werte, die der Menge S_k angehören. In der k-ten Gruppe seien n_k $(k = 1, 2, \ldots, r)$ Elemente vorhanden. In verschiedenen Stichproben vom Umfang n können für ein festes k die Werte n_k verschieden ausfallen; n_k ist für feste k eine binomial verteilte Zufallsvariable. Die Stichprobenfunktion

$$\chi^2 = \sum_{k=1}^{r} \frac{(n_k - n\pi_k)^2}{n\pi_k} \tag{12.4.2}$$

nennen wir die *Pearsonsche Stichprobenfunktion* χ^2.

Satz 12.4.1. *Die theoretischen Häufigkeiten π_k seien bestimmt. Die Folge $\{F_n(z)\}$ der Verteilungsfunktionen der durch die Formel* (12.4.2) *bestimmten Stichprobenfunktion χ^2 genügt dann den Beziehungen*

$$\lim_{n\to\infty} F_n(z) = \begin{cases} \dfrac{1}{2^{\frac{r-1}{2}}\,\Gamma\left(\dfrac{r-1}{2}\right)} \displaystyle\int_0^z z^{\frac{r-3}{2}} e^{-\frac{z}{2}}\, dz & \textit{für } z > 0, \\ 0 & \textit{für } z \leqq 0. \end{cases} \tag{12.4.3}$$

Der Ausdruck (12.4.3) stellt eine χ^2-Verteilung (siehe 9.4) mit $r-1$ Freiheitsgraden dar.

Beweis. Die Wahrscheinlichkeit dafür, daß unter den Bedingungen dieses Satzes n_1 Beobachtungen zur Menge S_1, n_2 zur Menge $S_2, \ldots, n_r$ zur Menge S_r gehören, wobei $n_1 + n_2 + \cdots + n_r = n$ gilt, ist gleich

$$\frac{n!}{n_1!\,n_2!\cdots n_r!}\,\pi_1^{n_1}\pi_2^{n_2}\cdots\pi_r^{n_r}. \tag{12.4.4}$$

Dieser Ausdruck stellt die Wahrscheinlichkeit in einer Polynomialverteilung (siehe 5.12) dar. Die charakteristische Funktion $\varphi(t_1, \ldots, t_r)$ dieser durch (12.4.4) gegebenen Wahrscheinlichkeitsfunktion hat die Form

$$\varphi(t_1, t_2, \ldots, t_r) = \left(\sum_{k=1}^{r} \pi_k e^{it_k}\right)^n.$$

Wir betrachten die Zufallsvariable

$$Y_k = \frac{n_k - n\pi_k}{\sqrt{n\pi_k}} \qquad (k = 1, 2, \ldots, r).$$

Es ist

$$\chi^2 = \sum_{k=1}^{r} Y_k^2, \qquad \sum_{k=1}^{r} Y_k\sqrt{\pi_k} = 0.$$

Aus der letzten Gleichung ersehen wir, daß die Variablen Y_k linear abhängig sind. Die charakteristische Funktion der gemeinsamen Verteilung von $(Y_1, Y_2, \ldots, Y_r)$ hat die Form

$$\varphi(t_1, t_2, \ldots, t_r) = \exp\left(-\sum_{k=1}^{r} it_k\sqrt{n\pi_k}\right)\left(\sum_{k=1}^{r}\pi_k \exp\left(\frac{it_k}{\sqrt{n\pi_k}}\right)\right)^n. \tag{12.4.5}$$

Daraus erhalten wir

$$\log \varphi(t_1, t_2, \ldots, t_r) = -i\sqrt{n}\sum_{k=1}^{r} t_k\sqrt{\pi_k} + n\log\left(\sum_{k=1}^{r}\pi_k \exp\left(\frac{it_k}{\sqrt{n\pi_k}}\right)\right)$$

$$= -i\sqrt{n}\sum_{k=1}^{r} t_k\sqrt{\pi_k} + n\log\left(1 + \frac{i}{\sqrt{n}}\sum_{k=1}^{r} t_k\sqrt{\pi_k} - \frac{1}{2n}\sum_{k=1}^{r} t_k^2 + o\left(\frac{1}{n}\right)\right)$$

$$= -i\sqrt{n}\sum_{k=1}^{r} t_k\sqrt{\pi_k} + n\log(1+z)$$

mit

$$z = \frac{i}{\sqrt{n}}\sum_{k=1}^{r} t_k\sqrt{\pi_k} - \frac{1}{2n}\sum_{k=1}^{r} t_k^2 + o\left(\frac{1}{n}\right).$$

Wenn wir beachten, daß z für $n \to \infty$ gegen Null strebt, so können wir $\log(1+z)$ in eine Reihe entwickeln. Wir erhalten so

$$n\log(1+z) - i\sqrt{n}\sum_{k=1}^{r} t_k\sqrt{\pi_k} \to -\frac{1}{2}\left[\sum_{k=1}^{r} t_k^2 - \left(\sum_{k=1}^{r} t_k\sqrt{\pi_k}\right)^2\right].$$

Schließlich ergibt sich

$$\lim_{n\to\infty} \varphi(t_1, t_2, \ldots, t_r) = \exp\left\{-\frac{1}{2}\left[\sum_{k=1}^{r} t_k^2 - \left(\sum_{k=1}^{r} t_k\sqrt{\pi_k}\right)^2\right]\right\}. \tag{12.4.6}$$

Wir zeigen nun, daß man eine orthogonale Transformation

$$u_k = \sum_{j=1}^{r} a_{kj} t_j \qquad (k = 1, 2, \ldots, r)$$

finden kann, deren Koeffizienten a_{rj} durch

$$a_{rj} = \sqrt{\pi_j} \qquad (j = 1, 2, \ldots, r)$$

gegeben sind.

Hier sind r^2 Koeffizienten unbekannt. Aus der Orthogonalitätsbedingung folgt, daß diese die $r + \dfrac{r(r-1)}{2}$ Gleichungen

$$\sum_{j=1}^{r} a_{kj} a_{ij} = \begin{cases} 1 & \text{für } k = i, \\ 0 & \text{für } k \neq i \end{cases}$$

erfüllen müssen. Überdies legten wir $r - 1$ Koeffizienten fest (in Wirklichkeit bestimmten wir r Koeffizienten, die aber durch die Beziehung $\sum_{j=1}^{r} a_{rj}^2 = 1$ mit-

einander verknüpft sind). Es bleiben also noch $r^2 - (r-1)$ Koeffizienten unbestimmt; diese können wir aber bestimmen, denn es ist

$$r^2 - (r-1) - r - \frac{r(r-1)}{2} = \frac{r^2 - 3r + 2}{2} \geqq 0 \quad \text{für} \quad r \geqq 2.$$

Die quadratische Form, die auf der rechten Seite der Formel (12.4.6) vorkommt, nimmt nach Einsetzen der Werte $u_1, u_2, \ldots, u_r$ die folgende Gestalt an:

$$\sum_{k=1}^{r-1} u_k^2.$$

Der Formel (12.4.6) wegen erhalten wir, daß die charakteristische Funktion der gemeinsamen Zufallsvariablen $(Y_1, Y_2, \ldots, Y_r)$ für $n \to \infty$ gegen die charakteristische Funktion der gemeinsamen Verteilung von $r-1$ unabhängigen Zufallsvariablen strebt, deren jede nach $N(0;1)$ verteilt ist. Aus einem in diesem Buch nicht bewiesenen Satz, der eine Verallgemeinerung von Satz 6.6.1b auf mehrdimensionale Verteilungen darstellt, folgt, daß $(Y_1, Y_2, \ldots, Y_r)$ eine asymptotische Normalverteilung hat, die durch den Limes der charakteristischen Funktionen bestimmt ist. Schließlich erhalten wir also, daß die durch Formel (12.4.2) bestimmte Stichprobenfunktion χ^2 asymptotisch eine Quadratsumme von $r-1$ unabhängigen Zufallsvariablen ist, die nach $N(0;1)$ verteilt sind, und folglich als Grenzverteilung eine χ^2-Verteilung mit $r-1$ Freiheitsgraden hat. Damit ist der Satz bewiesen.

C. Satz 12.4.1 bildet die Grundlage eines weitverbreiteten Anpassungstests, der χ^2-*Test* genannt wird. Wir erläutern seine Bedeutung durch ein Beispiel.

Beispiel 12.4.1. Einer Warenladung entnehmen wir eine $n = 100$ Elemente zählende einfache Stichprobe und finden in ihr $n_1 = 22$ fehlerhafte Stücke. Auf Grund dieser Stichprobe wollen wir mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}01$ die Hypothese H_0 testen, die aussagt, daß sich während der Stichprobenerhebung die Wahrscheinlichkeit dafür, ein fehlerhaftes Exemplar auszulosen, nicht änderte und stets $p = 0{,}20$ betrug. Mit anderen Worten, nach dieser Hypothese ist $n_1 = 22$ der beobachtete Wert einer binomial verteilten Zufallsvariablen mit $p = 0{,}20$ und $n = 100$.

Wir benutzen den χ^2-Test. Durch die Hypothese H_0 ist die Verteilung vollständig bestimmt, und wir können den Satz 12.4.1 anwenden.

Die Stichprobe teilen wir in n_1 fehlerhafte und $n_2 = n - n_1$ fehlerlose Stücke auf. Die theoretischen Häufigkeiten dieser Gruppen betragen $p = 0{,}2$ bzw. $q = 1 - p = 0{,}8$. Aus Formel (12.4.2) erhalten wir

$$\chi^2 = \frac{(n_1 - np)^2}{np} + \frac{(n_2 - nq)^2}{nq} = \frac{2^2}{20} + \frac{2^2}{80} = \frac{1}{4}.$$

Hier haben wir zwei Gruppen, also $r = 2$, und wir betrachten daher die χ^2-Verteilung mit einem Freiheitsgrad. Aus den χ^2-Tafeln lesen wir $P(\chi^2 \geqq 6{,}6) = 0{,}01$ ab; also ist die Wahrscheinlichkeit, daß der Wert χ^2 nicht kleiner als $\frac{1}{4}$ ist, bedeutend größer als 0,01. Es besteht also kein Grund, die Hypothese H_0 abzulehnen.

Bei Anwendung des χ^2-Tests ist zu bedenken, daß dieser auf dem Satz über die Grenzverteilung beruht. Die in diesem Test auftretenden Zahlen n_k können daher nicht klein sein. Man kann hier zur Orientierung die Regel angeben, derzufolge die Anzahl der Beobachtungen so in Gruppen zu zerlegen ist, daß $n \pi_k \geqq 10$ ist, d. h., daß die mittlere Anzahl der Beobachtungen in jeder Gruppe nicht kleiner ist als 10. Manche Autoren nehmen als Regel an, daß in jeder Gruppe nicht weniger als fünf Elemente auftreten; dies bezieht sich speziell auf Randgruppen. Einer tieferen Untersuchung des Problems der Anzahl der Gruppen, in die sämtliche Beobachtungen zu unterteilen sind, ist die Arbeit von MANN und WALD [1] gewidmet. In dieser Arbeit empfehlen MANN und WALD die Zerlegung der x-Achse in Intervalle mit gleichen theoretischen Häufigkeiten. Ist dann die Anzahl der Intervalle gleich r, so ist demzufolge $\pi_k = \frac{1}{r}$ $(k = 1, 2, \ldots, r)$ zu setzen. Im Zusammenhang mit der Arbeit von MANN und WALD gibt WILLIAMS [1] numerische Daten zur Anzahl und zur Länge dieser Intervalle an. Eine ausführliche Beschreibung der Anwendbarkeit des χ^2-Tests findet man in der Arbeit von COCHRAN [2]. Es ist erwähnenswert, daß VESSEREAU [1] durch Vergleich der exakten, durch (12.4.2) definierten χ^2-Verteilung mit der χ^2-Grenzverteilung mit $r - 1$ Freiheitsgraden für $\pi_k = \frac{1}{r}$ und kleine n (etwa 15 bis 20) festgestellt hat, daß man, ohne einen größeren Fehler zu befürchten, die χ^2-Grenzverteilung für den χ^2-Test mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ oder $\alpha = 0{,}01$ anwenden kann, sogar dann, wenn $n_k = 1$ $(k = 1, 2, \ldots, r)$ ist.

TUMANJAN [1] und GICHMAN [4] untersuchten die χ^2-Grenzverteilung für den Fall, daß die Anzahl der Intervalle, in die die $(-\infty, \infty)$-Achse zerlegt wird, für $n \to \infty$ gegen ∞ geht (siehe Aufgabe 12.8.7).

D. Bis jetzt betrachten wir nur den Fall, daß die Hypothese H_0 die Verteilung vollständig bestimmt. Wenn dagegen die Hypothese H_0 nur die Gestalt der Verteilung, nicht aber die Parameter $\lambda_1, \lambda_2, \ldots, \lambda_m$ bestimmt, so kann man auch nicht die Werte von π_k [nach der Formel (12.4.1)] ermitteln. In diesem Fall kann der Satz 12.4.1 nicht mehr als Grundlage für einen Test dienen.

FISHER [4] zeigte jedoch, daß man in diesem Fall den Satz 12.4.1 nur etwas zu ändern braucht, um ihn wieder anwenden zu können. Die Darlegung des Fisherschen Ergebnisses erfordert die Einführung des Begriffs der Likelihood-Funktion. Die Bedeutung dieser Funktion werden wir im nächsten Kapitel genauer kennenlernen.

Wir nehmen an, die Gestalt der Verteilungsfunktion $F(x)$, nach der ein Merkmal X in einer gewissen Gesamtheit verteilt ist, sei bekannt, während die Parameter $\lambda_1, \lambda_2, \ldots, \lambda_m$ unbekannt sind. Der Gesamtheit entnehmen wir eine einfache Stichprobe vom Umfang n, wobei die Werte $x_1, x_2, \ldots, x_n$ beobachtet werden.

Wir unterscheiden zwei Fälle, je nachdem, ob die Zufallsvariable X diskret oder stetig ist.

Nehmen wir an, die Zufallsvariable X sei diskret. Es bezeichne p_k die Wahrscheinlichkeit für das Eintreffen von $X = x_k$ $(k = 1, 2, \ldots, n)$. Die Wahrscheinlichkeiten p_k sind Funktionen der unbekannten Parameter $\lambda_1, \lambda_2, \ldots, \lambda_m$.

Definition 12.4.1. Für eine diskrete Zufallsvariable bezeichnen wir das Produkt

$$L = p_1(\lambda_1, \lambda_2, \ldots, \lambda_m)\, p_2(\lambda_1, \lambda_2, \ldots, \lambda_m) \cdots p_n(\lambda_1, \lambda_2, \ldots, \lambda_m) \tag{12.4.7}$$

als *Likelihood-Funktion.*

Nun sei die Zufallsvariable X stetig verteilt mit der Dichte $f(x, \lambda_1, \lambda_2, \ldots, \lambda_m)$.

Definition 12.4.2. Für stetige Zufallsvariable bezeichnen wir das Produkt

$$L = f(x_1, \lambda_1, \lambda_2, \ldots, \lambda_m) f(x_2, \lambda_1, \lambda_2, \ldots, \lambda_m) \cdots f(x_n, \lambda_1, \lambda_2, \ldots, \lambda_m) \tag{12.4.7'}$$

als *Likelihood-Funktion.*

Natürlich hängt die Likelihood-Funktion L von den Parametern

$$\lambda_1, \lambda_2, \ldots, \lambda_m$$

ab. Wenn wir die Werte dieser m unbekannten Parameter aus dem System der m Gleichungen

$$\frac{\partial \log L}{\partial \lambda_i} = 0 \qquad (i = 1, 2, \ldots, m) \tag{12.4.8}$$

bestimmen, so sagen wir, daß wir das Lösungssystem $\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*$ dieser Gleichungen nach dem *Maximum-Likelihood-Prinzip* gewonnen haben.

Wir wollen nun annehmen, wir hätten die x-Achse in r Intervalle zerlegt und entsprechend die n Beobachtungen in r Gruppen zusammengefaßt. Es mögen π_k und n_k die gleiche Bedeutung haben wie in (12.4.2). Offenbar sind die π_k Funktionen der unbekannten Parameter $\lambda_1, \lambda_2, \ldots, \lambda_m$. Wir nehmen $m < r$ an und betrachten den Ausdruck

$$L = \pi_1^{n_1}(\lambda_1, \lambda_2, \ldots, \lambda_m) \cdots \pi_r^{n_r}(\lambda_1, \lambda_2, \ldots, \lambda_m). \tag{12.4.9}$$

Dieser Ausdruck ist die Likelihood-Funktion *nach der Gruppierung der Beobachtungen.* Der früher erwähnte Fishersche Satz, der von Cramér ([2], § 30,3) präzisiert wurde, lautet:

Satz 12.4.2. *Es seien alle* $\pi_k(\lambda_1, \lambda_2, \ldots, \lambda_m) > 0$, *und es mögen die stetigen Ableitungen* $\dfrac{\partial \pi_k}{\partial \lambda_i}$, $\dfrac{\partial^2 \pi_k}{\partial \lambda_i \partial \lambda_j}$ $(k = 1, 2, \ldots, r;\ i,\ j = 1, 2, \ldots, m)$ *existieren. Ferner habe die Matrix der* $\dfrac{\partial \pi_k}{\partial \lambda_i}$ $(k = 1, 2, \ldots, r;\ i = 1, 2, \ldots, m)$ *den Rang* m. *Sind dann die unbekannten Parameter* $\lambda_1, \lambda_2, \ldots, \lambda_m$ *nach dem Maximum-Likelihood-*

Prinzip nach Gruppierung der Beobachtungen, d. h. aus dem Gleichungssystem (12.4.8) *ermittelt worden, wobei L durch* (12.4.9) *gegeben ist, so strebt die Verteilung der durch* (12.4.2) *definierten Stichprobenfunktion* χ^2 *gegen die* χ^2-*Verteilung mit* $r - m - 1$ *Freiheitsgraden.*

Den Beweis dieses Satzes findet der Leser bei CRAMÉR [2].

Die Auflösung des Gleichungssystems (12.4.8) stößt, wenn L durch (12.4.9) gegeben ist, in der Praxis auf große Schwierigkeiten. Gewöhnlich ist es weitaus günstiger, das Gleichungssystem (12.4.8) aufzulösen, wenn die Likelihood-Funktion L durch (12.4.7) oder durch (12.4.7′) gegeben ist, d. h., wenn sich die Likelihood-Funktion aus den Beobachtungen vor ihrer Gruppierung errechnet. Jedoch ist dann der Satz 12.4.2 nicht mehr erfüllt. Darauf hat FISHER [9] hingewiesen. CHERNOFF und LEHMANN [1] haben folgenden Sachverhalt nachgewiesen: Genügen die Funktionen $\pi_k(\lambda_1, \lambda_2, \ldots, \lambda_m)$ den Voraussetzungen des Satzes 12.4.2 und genügt die Dichtefunktion $f(x, \lambda_1, \ldots, \lambda_m)$ oder evtl. die Wahrscheinlichkeitsfunktion $p(\lambda_1, \lambda_2, \ldots, \lambda_m)$ den Voraussetzungen des Satzes 13.7.3 oder besser noch den Voraussetzungen der Verallgemeinerung dieses Satzes auf den Fall mehrerer unbekannter Parameter (siehe 13.7), so hat die durch (12.4.2) definierte Stichprobenfunktion χ^2 für $n \to \infty$ die gleiche Verteilung wie $U + a_1 Y_1^2 + \cdots + a_m Y_m^2$, wobei U die χ^2-Verteilung mit $r - m - 1$ Freiheitsgraden hat, während die Y_1, $Y_2, \ldots, Y_m$ die Normalverteilung $N(0;1)$ haben. Dabei sind die U, $Y_1, Y_2, \ldots, Y_m$ unabhängig, während die $a_1, a_2, \ldots, a_m$ gewisse Zahlen zwischen 0 und 1 sind (das Verfahren zur Berechnung der $a_1, a_2, \ldots, a_m$ ist bei CHERNOFF und LEHMANN [1] angegeben). Ist die Anzahl r der Gruppen groß, so kann man ohne Bedenken die Behauptung des Satzes 12.4.2 anwenden, wobei L nach (12.4.7) oder nach (12.4.7′) zu berechnen ist. Ist hingegen r klein, so sind die Ergebnisse, die man aus einer derartigen Anwendung des Satzes 12.4.2 gewinnt, mit äußerster Vorsicht in Anwendung zu bringen.

WATSON [1], [2] kommt bei der Betrachtung des Falles, daß die x-Achse in Intervalle mit konstanten theoretischen Häufigkeiten $\pi_k = \dfrac{1}{r}$ zerlegt wird, zu Folgerungen, die denen von CHERNOFF und LEHMANN [1] analog sind. Wir empfehlen dem Leser die Arbeiten von NEYMAN und PEARSON [1] sowie von NEYMAN [8]. In der zuletzt genannten Arbeit werden gewisse Modifikationen und Verallgemeinerungen des χ^2-Tests betrachtet. Darüber hinaus wird dort eine Methode zur Bestimmung der unbekannten Parameter aus der Stichprobe angegeben, dessen Spezialfall das Maximum-Likelihood-Prinzip ist.

Beispiel 12.4.2. Dieses Beispiel wurde dem Buch von BOJEW [1] entnommen. Man hat 300 Baumwollgarnrollen, die einer Baumwollpartie entnommen wurden, auf ihre Festigkeit untersucht. Mit X wurde die Festigkeit in Kilogramm bezeichnet. Die beobachteten Werte x_k wurden in 13 Gruppen aufgeteilt, wobei n_i die Anzahl der Beobachtungen, die zur i-ten Gruppe gerechnet wurden, bezeichnet. Die Angaben wurden in Tabelle 12.4.1 zusammengestellt.

Tabelle 12.4.1.

i	x	n_i	i	x	n_i
1	0,5—0,64	1	8	1,48—1,62	53
2	0,64—0,78	2	9	1,62—1,76	25
3	0,78—0,92	9	10	1,76—1,90	19
4	0,92—1,06	25	11	1,90—2,04	16
5	1,06—1,20	37	12	2,04—2,18	3
6	1,20—1,34	53	13	2,18—2,38	1
7	1,34—1,48	56			

Wir wollen hier die Hypothese H_0 testen, daß das Merkmal X normal verteilt ist. Durch diese Hypothese ist weder der Mittelwert noch die Standardabweichung bestimmt; denn H_0 besagt nur, daß X irgendeine Normalverteilung besitzt. Wir müssen also den Wert dieser Parameter aus der Stichprobe berechnen, und zwar mit Hilfe des Maximum-Likelihood-Prinzips.

Die Dichte der Zufallsvariablen X hat die Form

$$f(x, m, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-m)^2}{2\sigma^2}},$$

wobei die Parameter m und σ unbekannt sind. Die Likelihood-Funktion ist auf Grund von (12.4.7′) gleich

$$L = \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n \exp\left[-\frac{1}{2\sigma^2}\sum_{l=1}^{n}(x_l - m)^2\right] \qquad (n = 300).$$

Daraus ergibt sich

$$\log L = -\frac{1}{2\sigma^2}\sum_{l=1}^{n}(x_l - m)^2 - \frac{1}{2}\, n \log \sigma^2 - \frac{1}{2}\, n \log 2\pi.$$

Wenden wir die Formel (12.4.8) an, so erhalten wir

$$\frac{\partial \log L}{\partial m} = \frac{1}{\sigma^2}\sum_{l=1}^{n}(x_l - m) = 0,$$

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{1}{2\sigma^4}\sum_{l=1}^{n}(x_l - m)^2 - \frac{n}{2\sigma^2} = 0. \tag{12.4.10}$$

Die Auflösung dieses Gleichungssystems ergibt

$$m^* = \frac{1}{n}\sum_{l=1}^{n} x_l = \bar{x},$$

$$\sigma^{*2} = \frac{1}{n}\sum_{l=1}^{n}(x_l - \bar{x})^2 = s^2. \tag{12.4.11}$$

Also sind der Stichprobenmittelwert und die Stichprobendispersion gleichzeitig die Werte m^* und σ^*, die man nach dem Maximum-Likelihood-Prinzip erhält.

Da die Intervalle in Tabelle 12.4.1 ziemlich schmal sind, können wir annehmen, daß die Beobachtungen in dem Mittelpunkt x_i eines jeden Intervalls liegen. Aus Tabelle 12.4.1 erhalten wir auf diese Weise Tabelle 12.4.2.

Tabelle 12.4.2.

i	x_i	n_i	i	x_i	n_i
1	0,57	1	8	1,55	53
2	0,71	2	9	1,69	25
3	0,85	9	10	1,83	19
4	0,99	25	11	1,97	16
5	1,13	37	12	2,11	3
6	1,27	53	13	2,28	1
7	1,41	56			

Die Berechnung ergibt

$$m^* = 1{,}41, \qquad \sigma^* = 0{,}26.$$

Da die Randgruppen in Tabelle 12.4.1 sehr klein sind, vereinigen wir die zwei tiefsten und die zwei höchsten Gruppen in je eine. Um die theoretischen Häufigkeiten π_k zu berechnen, benutzen wir die Tafeln der Normalverteilung. Die Intervallenden legen wir in die Mitte des Abstandes zwischen den benachbarten Punkten x_i und x_{i+1}. Für eine Veränderliche X mit der Verteilung $N(1{,}41; 0{,}26)$ ergibt sich

$$\pi_1 = P(X < 0{,}78) + P(X \geqq 2{,}04) = P\left(\frac{X - 1{,}41}{0{,}26} < -2{,}42\right) + P\left(\frac{X - 1{,}41}{0{,}26} \geqq 2{,}42\right) = 0{,}0156.$$

$$\pi_2 = P(0{,}78 \leqq X < 0{,}92) = P\left(-2{,}42 \leqq \frac{X - 1{,}41}{0{,}26} < -1{,}88\right) = 0{,}0223,$$

$$\pi_3 = P(0{,}92 \leqq X < 1{,}06) = P\left(-1{,}88 \leqq \frac{X - 1{,}41}{0{,}26} < -1{,}35\right) = 0{,}0584,$$

$$\pi_4 = P(1{,}06 \leqq X < 1{,}20) = P\left(-1{,}35 \leqq \frac{X - 1{,}41}{0{,}26} < -0{,}81\right) = 0{,}1205,$$

$$\pi_5 = P(1{,}20 \leqq X < 1{,}34) = P\left(-0{,}81 \leqq \frac{X - 1{,}41}{0{,}26} < -0{,}27\right) = 0{,}1846,$$

$$\pi_6 = P(1{,}34 \leqq X < 1{,}48) = P\left(-0{,}27 \leqq \frac{X - 1{,}41}{0{,}26} < 0{,}27\right) = 0{,}2128,$$

$$\pi_7 = P(1{,}48 \leqq X < 1{,}62) = P\left(0{,}27 \leqq \frac{X - 1{,}41}{0{,}26} < 0{,}81\right) = 0{,}1846,$$

$$\pi_8 = P(1{,}62 \leqq X < 1{,}76) = P\left(0{,}81 \leqq \frac{X - 1{,}41}{0{,}26} < 1{,}35\right) = 0{,}1205,$$

$$\pi_9 = P(1{,}76 \leqq X < 1{,}90) = P\left(1{,}35 \leqq \frac{X - 1{,}41}{0{,}26} < 1{,}88\right) = 0{,}0584,$$

$$\pi_{10} = P(1{,}90 \leqq X < 2{,}04) = P\left(1{,}88 \leqq \frac{X - 1{,}41}{0{,}26} < 2{,}42\right) = 0{,}0223.$$

Wir berechnen

$$\chi^2 = \sum_{k=1}^{10} \frac{(n_k - n\pi_k)^2}{n\pi_k} = 22{,}07.$$

Da wir zwei Parameter aus der Stichprobe bestimmt haben, beträgt hier die Anzahl der Freiheitsgrade 7. Aus den Tafeln der χ^2-Verteilung finden wir, daß $P(\chi^2 \geqq 22{,}07) < 0{,}01$ ist. Wir lehnen also die Hypothese H_0 ab.

Beispiel 12.4.3. Im Beispiel 5.5.1 stellten wir eine gute Übereinstimmung zwischen der empirischen Verteilung der Todesfälle, die durch Huftritte verursacht wurden, und der Poissonschen Verteilung fest. Dabei wählten wir als Parameterwert λ den Mittelwert der empirischen Verteilung.

Im Grunde genommen wurde hier die Hypothese aufgestellt, daß die betrachtete Zufallsvariable eine Poissonsche Verteilung hat, wobei, ohne eine Sicherheitswahrscheinlichkeit festzulegen, angenommen wurde, daß die Beobachtungsergebnisse mit der Hypothese übereinstimmen.

Jetzt wollen wir einen χ^2-Test mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ anwenden, um festzustellen, ob die beobachtete Divergenz zwischen der empirischen und der hypothetischen Verteilung nicht gegen die angenommene Hypothese spricht.

Wir ändern die Tabelle 5.5.1, indem wir statt der Häufigkeit und Wahrscheinlichkeiten die beobachteten und die erwarteten Besetzungszahlen angeben, und erhalten so die Tabelle 12.4.3.

Tabelle 12.4.3.

i	0	1	2	3	4
beobachtete Anzahl n_i	109	65	22	3	1
erwartete Anzahl np_i	108,8	66,2	20,2	4,2	0,6

Die beiden letzten Gruppen hat man in eine vereint, da ihre Besetzungszahlen sehr klein sind. Es gilt

$$\chi^2 = \sum_{i=1}^{4} \frac{(n_i - np_i)^2}{np_i} = 0{,}3160.$$

Da aus den Beobachtungen ein Parameter[1]) errechnet wurde, hat die Stichprobenfunktion χ^2 zwei Freiheitsgrade. Den Tafeln für die χ^2-Verteilung entnehmen wir, daß der Wert $P(\chi^2 \geqq 0{,}3160)$ viel größer als $\alpha = 0{,}05$ ist. Es besteht also kein Grund, die zu untersuchende Hypothese abzulehnen.

12.5. Anpassungstests, die sich auf die Sätze von Kolmogoroff und Smirnow stützen

A. Der Satz 10.11.1 von KOLMOGOROFF kann ebenfalls als Grundlage für einen Anpassungstest dienen, der die Hypothese H_0, durch welche eine stetige Verteilungsfunktion $F(x)$ festgelegt wird, verifizieren soll. Wir wollen hier diesen Satz noch einmal formulieren.

Es sei $S_n(x)$ die empirische Verteilungsfunktion (d. h. die Häufigkeit des Ereignisses $X < x$) in einer einfachen Stichprobe vom Umfang n. Weiter setzen wir

$$D_n = \sup_{-\infty < x < \infty} |F(x) - S_n(x)|. \tag{12.5.1}$$

Nach dem Kolmogoroffschen Satz besteht für $\lambda > 0$ die Gleichung

$$\lim_{n\to\infty} Q_n(\lambda) = \lim_{n\to\infty} P\left(D_n < \frac{\lambda}{\sqrt{n}}\right) = Q(\lambda) \tag{12.5.2}$$

mit

$$Q(\lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}. \tag{12.5.3}$$

Wir nehmen als Hypothese H_0 an, daß das Merkmal X in der Gesamtheit nach der stetigen Verteilungsfunktion $F(x)$ verteilt sei. Dieser Gesamtheit entnehmen wir eine genügend umfangreiche einfache Stichprobe und berechnen die empirische Verteilungsfunktion $S_n(x)$ und die Stichprobenfunktion D_n.

Um die Hypothese mit der Sicherheitswahrscheinlichkeit α nachzuprüfen, suchen wir in Tafel VIII einen solchen Wert λ_0, daß $Q(\lambda_0) = 1 - \alpha$ ist. Ist der beobachtete Wert von D_n größer oder gleich $\frac{\lambda_0}{\sqrt{n}}$, so lehnen wir die Hypothese H_0 ab. Im entgegengesetzten Fall besteht kein Grund zu ihrer Ablehnung.

Beispiel 12.5.1. In Tabelle 12.5.1 ist auf Grund einer Arbeit von WISZNIEWSKI [1] die Verteilung der Monatsdurchschnittstemperaturen in Warschau für die Jahre von 1779 bis 1947 (ohne das Jahr 1945) angegeben. Die Gesamtanzahl der Beobachtungen beträgt also $n = 168$.

---

[1]) Im Beispiel 13.7.1 werden wir beweisen, daß $\bar{x}$ die Abschätzung des unbekannten Parameters λ nach dem Maximum-Likelihood-Prinzip darstellt.

Es sei X die Durchschnittstemperatur für Januar in Warschau; mit x_k bzw. n_k ist in Tabelle 12.5.1 der beobachtete Wert der Zufallsvariablen X bzw. die Anzahl der Beobachtungen x_k bezeichnet. Aus dieser Tabelle hat man

$$\bar{x} = -4{,}22, \qquad s = 3{,}57$$

berechnet, wobei $\bar{x}$ bzw. s der Stichprobenmittelwert bzw. die Stichprobenstandardabweichung sind. Wir stellen die Hypothese H_0 auf, wonach die Zufallsvariable X die Normalverteilung $N(-4{,}22; 3{,}57)$ hat, und verifizieren diese Hypothese mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$. Dazu benutzen wir den Satz von KOLMOGOROFF. Wir bestimmen die Werte der Verteilungsfunktion einer Zufallsvariablen mit $N(-4{,}22; 3{,}57)$ sowie die Werte der empirischen Verteilungsfunktion in den Punkten x_k. Diese Werte sind in Tabelle 12.5.1 zusammengestellt. Der beobachtete Wert d_n der Zufallsvariablen D_n ist gleich dem größten Wert der beobachteten Differenzen

$$|s_n(x_{k+1}) - F(x_k)| \quad \text{und} \quad |s_n(x_k) - F(x_k)|.$$

Wir erhalten also

$$d_n = |s_n(-5{,}4) - F(-5{,}4)| = 0{,}086$$

und mit Hilfe von Tafel VIII

$$P(D_n \geqq 0{,}086) = P\left(D_n \sqrt{168} \geqq 1{,}115\right) \approx 0{,}1663.$$

Dieser Test führte also nicht zur Ablehnung der Hypothese H_0.

Wir machen den Leser darauf aufmerksam, daß der in diesem Beispiel angewandte Gedankengang nicht vollkommen exakt ist. GICHMAN [2] hat nämlich gezeigt, daß der Satz 10.11.1 nicht erfüllt zu sein braucht, wenn die theoretische Verteilungsfunktion von unbekannten Parametern abhängig ist, welche durch Stichproben abgeschätzt werden. Die Ergebnisse, die man in dem betrachteten Fall durch Anwendung des Kolmogoroffschen Satzes erhält, sind also mit Vorsicht zu betrachten.

Im Zusammenhang mit Beispiel 12.5.1 drängt sich folgende Bemerkung auf. In der Praxis beobachtet man die untersuchten Werte nur mit einer gewissen Genauigkeit, die beispielsweise durch die angenommenen Maßeinheiten bedingt ist. So hat man in Beispiel 12.5.1 die Temperatur mit einer Genauigkeit bis zu 0,1°C gemessen. Infolgedessen wurden die Beobachtungen, die im Intervall $(x_k - 0{,}05°\text{C}, x_k + 0{,}05°\text{C})$ liegen, im Punkt x_k gruppiert. Wie GICHMAN [1] zeigte, hat die Gruppierung von Beobachtungen einen Einfluß auf den Grenzwert der Folge $\{Q_n(\lambda)\}$. GICHMAN hat aber festgestellt, daß man den Kolmogoroffschen Satz anwenden kann, wenn die Länge der Gruppierungsintervalle für $n \to \infty$ gleichmäßig gegen Null konvergiert. Man kann also den Kolmogoroffschen Satz dann anwenden, wenn alle Gruppierungsintervalle klein sind. Diese Bemerkung gilt auch für den Smirnowschen Satz.

B. Wir wollen jetzt einen Anpassungstest angeben, der sich auf den Satz 10.11.2 von SMIRNOW stützt. Dieser Satz kann als Grundlage für einen Test dienen, der

Tabelle 12.5.1.

x_k	n_k	$s_n(x_k)$	$F(x_k)$	x_k	n_k	$s_n(x_k)$	$F(x_k)$
−17,8	1	0,000	0,000	−4,2	1	0,440	0,504
−14,1	1	0,006	0,003	−4,1	3	0,446	0,512
−13,4	1	0,012	0,005	−4,0	2	0,464	0,524
−13,2	1	0,018	0,006	−3,9	6	0,476	0,536
−13,0	1	0,024	0,007	−3,8	2	0,511	0,548
−12,2	1	0,030	0,013	−3,6	3	0,523	0,567
−11,6	1	0,036	0,019	−3,5	2	0,541	0,579
−11,3	1	0,042	0,024	−3,4	1	0,553	0,591
−10,9	1	0,048	0,031	−3,3	2	0,559	0,603
−10,1	1	0,054	0,049	−3,2	3	0,571	0,614
−9,7	1	0,060	0,063	−3,1	1	0,589	0,622
−9,4	2	0,065	0,074	−3,0	3	0,595	0,633
−9,3	3	0,077	0,078	−2,9	4	0,612	0,644
−9,2	2	0,095	0,082	−2,8	4	0,636	0,655
−9,1	1	0,107	0,085	−2,6	1	0,660	0,674
−9,0	1	0,113	0,090	−2,5	1	0,666	0,684
−8,7	1	0,119	0,106	−2,4	3	0,672	0,695
−8,6	1	0,125	0,109	−2,3	2	0,690	0,705
−8,0	1	0,131	0,145	−2,1	1	0,702	0,722
−7,8	3	0,136	0,159	−2,0	4	0,708	0,732
−7,7	1	0,154	0,166	−1,9	2	0,731	0,742
−7,4	3	0,160	0,187	−1,7	3	0,743	0,761
−7,3	2	0,178	0,195	−1,6	2	0,761	0,767
−7,2	1	0,190	0,203	−1,4	1	0,773	0,785
−7,1	1	0,196	0,209	−1,3	1	0,779	0,794
−7,0	1	0,202	0,218	−1,2	4	0,785	0,802
−6,8	2	0,208	0,236	−1,1	3	0,809	0,808
−6,7	2	0,220	0,245	−1,0	3	0,827	0,816
−6,6	1	0,232	0,251	−0,8	2	0,844	0,832
−6,4	2	0,238	0,271	−0,6	1	0,856	0,844
−6,3	1	0,250	0,281	−0,3	1	0,862	0,864
−6,1	1	0,255	0,298	−0,2	4	0,868	0,871
−6,0	2	0,261	0,308	−0,1	2	0,892	0,875
−5,9	2	0,273	0,319	0,0	1	0,904	0,881
−5,4	2	0,285	0,371	0,1	1	0,910	0,887
−5,3	4	0,297	0,382	0,3	1	0,916	0,898
−5,2	2	0,321	0,394	0,6	1	0,922	0,911
−5,1	2	0,333	0,401	0,7	3	0,928	0,916
−5,0	1	0,345	0,413	0,9	1	0,946	0,924
−4,9	2	0,351	0,425	1,2	1	0,952	0,936
−4,8	1	0,363	0,436	1,3	2	0,958	0,939
−4,7	2	0,369	0,448	1,4	1	0,969	0,942
−4,6	2	0,380	0,456	1,6	1	0,975	0,948
−4,5	1	0,392	0,468	1,9	1	0,981	0,956
−4,4	4	0,398	0,480	2,6	1	0,987	0,972
−4,3	3	0,422	0,492	2,8	1	0,993	0,976

eine Hypothese H_0 nachprüft, daß zwei Proben aus Grundgesamtheiten stammen, deren Merkmal X dieselbe stetige Verteilungsfunktion besitzt. Wir betonen, daß in der Hypothese H_0 die Verteilungsfunktion der Gesamtheiten *nicht* präzisiert wird. Mit $S_{1n_1}(x)$ bzw. $S_{2n_2}(x)$ seien die empirischen Verteilungsfunktionen der unabhängigen einfachen Stichproben vom Umfang n_1 bzw. n_2 bezeichnet. Wenn diese Stichproben aus gleichverteilten Gesamtheiten stammen, dann strebt nach dem Satz von SMIRNOW die Folge der Verteilungsfunktionen $Q_{n_1n_2}(\lambda)$, die durch

$$Q_{n_1n_2}(\lambda) = P\left(D_{n_1n_2} < \frac{\lambda}{\sqrt{n}}\right) \tag{12.5.4}$$

bestimmt sind, wobei $n = \dfrac{n_1 n_2}{n_1 + n_2}$ und

$$D_{n_1n_2} = \sup_{-\infty<x<\infty} |S_{1n_1}(x) - S_{2n_2}(x)| \tag{12.5.5}$$

gilt, für $\lambda > 0$ mit $n \to \infty$ gegen die durch die Formel (12.5.3) bestimmte Verteilungsfunktion $Q(\lambda)$.

Gegeben seien zwei einfache Stichproben, die aus zwei Gesamtheiten stammen. Wir stellen die Hypothese H_0 auf, daß das Merkmal X in beiden Gesamtheiten dieselbe stetige Verteilungsfunktion besitzt. Ist der beobachtete Wert der Stichprobenfunktion $D_{n_1n_2}$ größer oder gleich $\dfrac{\lambda_0}{\sqrt{n}}$, wobei $Q(\lambda_0) = 1 - \alpha$ ist, so lehnen wir die Hypothese H_0 ab.

Beispiel 12.5.2. Im Beispiel 12.3.4 betrachteten wir den durchschnittlichen Brotverbrauch pro Normalverbraucher in zwei Stichproben. In der einen befanden sich $n_1 = 65$ Arbeiterfamilien aus Warschau und Łódź, in der anderen $n_2 = 57$ Arbeiterfamilien aus dem oberschlesischen Industriegebiet. Wir haben festgestellt, daß bei Anwendung eines Parametertestes mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ kein Grund besteht, die Hypothese H_0 abzulehnen, nach welcher der durchschnittliche Roggenbrotverbrauch für Arbeiterfamilien in Warschau und Łódź sowie im oberschlesischen Industriegebiet der gleiche ist. Wir wollen jetzt die weit stärkere Hypothese nachprüfen, nach der nicht nur der durchschnittliche Verbrauch, sondern auch die Verteilung der Arbeiterfamilien hinsichtlich des Roggenbrotverbrauches in Warschau und Łódź sowie im oberschlesischen Industriegebiet die gleichen sind. Wir wollen also die Hypothese H_0 prüfen, nach welcher beide Stichproben aus Gesamtheiten mit gleicher Verteilung erhoben wurden.

In Tabelle 12.5.2 ist mit x_k der monatliche Brotverbrauch pro Normalverbraucher bezeichnet. Die Zahlen n_k bedeuten die Anzahl der Familien, deren Brotverbrauch x_k beträgt, während $s_{1n_1}(x_k)$ und $s_{2n_2}(x_k)$ die empirischen Verteilungswerte beider Stichproben in den Punkten x_k bedeuten. Der Tabelle 12.5.2 entnehmen wir

$$d_{n_1n_2} = \max_x |s_{1n_1}(x) - s_{2n_2}(x)| = 0{,}105.$$

Tabelle 12.5.2.

x_k	n_k Warschau Lódź	oberschles. Industr.-Geb.	$s_{1n_1}(x_k)$	$s_{2n_2}(x_k)$	$\lvert s_{1n_1}(x_k) - s_{2n_2}(x_k)\rvert$
1	0	4	0,000	0,000	0,000
2	0	2	0,000	0,070	0,070
3	3	1	0,000	0,105	0,105
4	2	1	0,046	0,123	0,077
5	3	2	0,077	0,140	0,063
6	5	4	0,123	0,175	0,052
7	7	5	0,200	0,246	0,046
8	5	6	0,308	0,333	0,025
9	9	7	0,384	0,439	0,055
10	8	2	0,523	0,561	0,038
11	5	3	0,645	0,596	0,049
12	5	3	0,723	0,649	0,074
13	0	1	0,800	0,702	0,098
14	1	4	0,800	0,719	0,081
15	2	0	0,815	0,789	0,026
16	1	4	0,846	0,789	0,057
17	6	2	0,861	0,860	0,001
18	1	0	0,954	0,895	0,059
20	2	1	0,969	0,895	0,074
22	0	1	1,000	0,912	0,088
23	0	3	1,000	0,930	0,070
24	0	1	1,000	0,982	0,018
>24	0	0	1,000	1,000	0,000

Wegen $n = \dfrac{n_1 n_2}{n_1 + n_2} = 30{,}37$ folgt aus Tafel VIII

$$P(D_{n_1 n_2} \geqq 0{,}105) = P(D_{n_1 n_2}\sqrt{30{,}37} \geqq 0{,}58) \approx 0{,}8896.$$

Der angewandte λ-Test gibt also keine Handhabe, die Hypothese abzulehnen.

Beispiel 12.5.3. Aus drei Gesamtheiten, in denen das Merkmal X stetige Verteilungsfunktionen hat, hat man je eine einfache Stichprobe zu $n_1 = 100$, $n_2 = 150$ und $n_3 = 200$ Elementen erhoben. Zweck dieser Probeuntersuchung ist es, die Hypothese H_0 nachzuprüfen, nach der X in allen drei Gesamtheiten dieselbe Verteilung hat. Für die Stichprobenfunktion $D_i(N_i, n_{i+1})$ hat man die durch die Formeln (10.13.14) bestimmten Werte

$$d_1(N_1, n_2) = 0{,}20,$$

$$d_2(N_2, n_3) = 0{,}17$$

beobachtet. Wir prüfen die Hypothese H_0 mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$. Wir haben

$$\sqrt{\frac{100 \cdot 150}{250}} \cdot 0{,}2 = 1{,}55, \qquad \sqrt{\frac{250 \cdot 200}{450}} \cdot 0{,}17 = 1{,}79.$$

Da 1,79 die größere der beobachteten Zahlen ist, erhalten wir nach Satz 10.13.2 für große Stichproben die Näherungsgleichung

$$P\left(\max_{1\leqq i\leqq 2}\sqrt{\frac{N_i n_{i+1}}{N_{i+1}}}\,D_i(N_i, n_{i+1})\geqq 1{,}79\right)\approx 1-[Q(1{,}79)]^2.$$

Auf Grund von Tafel VIII ist diese Wahrscheinlichkeit gleich 0,0066. Wir lehnen also die Hypothese H_0 ab.

12.6. Die Tests von Wald-Wolfowitz und Wilcoxon-Mann-Whitney

A. Wald und Wolfowitz [1] konstruierten ein scharfsinniges Prüfverfahren für die Hypothese, daß zwei Stichproben aus Gesamtheiten mit der gleichen Verteilung stammen. Dieser Test hat als Grundlage die Iterationstheorie, die im vorhergehenden Kapitel besprochen wurde.

Es seien x_{1l} $(l = 1, \ldots, n_1)$ und x_{2j} $(j = 1, \ldots, n_2)$ zwei unabhängige einfache Stichproben, die aus zwei Gesamtheiten ausgelost wurden, in denen das Merkmal X eine stetige Verteilungsfunktion besitzt. Wegen

$$P(X_{1l}-X_{1l'}=0)=P(X_{2j}-X_{2j'}=0)=P(X_{1l}-X_{2l}=0)=0$$

können wir annehmen, daß

$$x_{1l}\neq x_{1l'}\quad (l'\neq l),\quad x_{2j}\neq x_{2j'}\,(j'\neq j),\quad x_{1l}\neq x_{2j}\quad (l=1,\ldots,n_1;\quad j=1,\ldots,n_2)$$

gilt. Wir ordnen die Werte x_{1l} und x_{2j} der Größe nach an, d. h., wir bilden aus den Elementen x_{1l} und x_{2j} die Folge z_m $(m = 1, \ldots, n_1 + n_2)$, wobei

$$z_1 < z_2 < \cdots < z_{n_1+n_2} \tag{12.6.1}$$

ist. Danach ordnen wir denjenigen z_m, die aus der ersten Stichprobe stammen, die Zahl 0 und den übrigen z_m die Zahl 1 zu. So erhalten wir eine Null-Eins-Folge, die n_1 Nullen und n_2 Einsen enthält. Ist die Hypothese H_0, daß beide Stichproben aus Gesamtheiten mit der gleichen Verteilung stammen, richtig, so treten alle Gruppen, die aus n_1 Nullen und aus n_2 Einsen bestehen, mit der gleichen Wahrscheinlichkeit auf. Man kann also die Iterationstheorie verwenden, um die Hypothese H_0 zu testen.

Wir müssen uns noch überlegen, welche Schlüsse wir bei Anwendung der Iterationstheorie zu ziehen haben. Schon von vornherein neigen wir dazu, die Hypothese H_0 abzulehnen, wenn es sich zeigt, daß die Anzahl der beobachteten Iterationen im Verhältnis zur erwarteten Anzahl klein ist; es treten dann nämlich lange Iterationen auf, die Elemente der einzelnen Stichproben bilden große Häufungen, und das zeugt davon, daß das Merkmal X nicht in beiden Gesamtheiten gleich verteilt ist. Mit anderen Worten, je größer die Anzahl der Iterationen ist und je kürzer die einzelnen Iterationen sind, desto stärker bezeugt die Stich-

probe die Richtigkeit der Hypothese H_0. Der Anpassungstest besteht hier also darin, einen Wert k_0 zu finden, für den $P(K \leqq k_0) \leqq \alpha$ ist, wobei α die Sicherheitswahrscheinlichkeit bedeutet, und die zu verifizierende Hypothese H_0 abzulehnen, wenn das beobachtete k höchstens gleich k_0 ist.

Beispiel 12.6.1. Um die Hypothese H_0, daß das Merkmal X die gleiche stetige Verteilung in zwei Gesamtheiten besitzt, zu verifizieren, entnahm man diesen Gesamtheiten zwei unabhängige einfache Stichproben, jede vom Umfang 4. Man erhielt die Werte

$$x_{11} = 3, \quad x_{12} = 3{,}5, \quad x_{13} = 6{,}5, \quad x_{14} = 2{,}5,$$
$$x_{21} = 7, \quad x_{22} = 2{,}9, \quad x_{23} = 4{,}0, \quad x_{24} = 6{,}0.$$

Die z-Werte sind hier

$$z_1 = 2{,}5, \quad z_2 = 2{,}9, \quad z_3 = 3, \quad z_4 = 3{,}5,$$
$$z_5 = 4, \quad z_6 = 6, \quad z_7 = 6{,}5, \quad z_8 = 7.$$

Wir erhalten also die Null-Eins-Folge

$$0\ 1\ 0\ 0\ 1\ 1\ 0\ 1.$$

Wenn wir die Bezeichnungen aus Kapitel 11 benutzen, haben wir

$$k_{11} = 2, \quad k_{12} = 1, \quad k_{13} = k_{14} = 0,$$
$$k_{21} = 2, \quad k_{22} = 1, \quad k_{23} = k_{24} = 0,$$
$$k_1 = k_2 = 3, \quad k = 6.$$

Wir beachten, daß die Umfänge der einzelnen Stichproben hier keine Zufallsvariablen darstellen. Um die gesuchten Wahrscheinlichkeiten zu berechnen, benutzen wir die Formeln (11.3.12) und erhalten

$$P(K = 2 \mid N_1 = 4,\ N_2 = 4) = 0{,}029,$$
$$P(K = 3 \mid N_1 = 4,\ N_2 = 4) = 0{,}086.$$

Wenn wir die Hypothese H_0 mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ testen, finden wir, daß der kritische Wert hier $k_0 = 2$ beträgt. Da wir $k = 6$ beobachtet haben, besteht kein Grund, die Hypothese H_0 abzulehnen. Wir bemerken auch, daß nach Formel (11.4.4) der Mittelwert $E(K \mid N_1 = 4,\ N_2 = 4)$ gleich 5, also beinahe gleich dem beobachteten Wert von k ist.

B. Ein anderer nichtparametrischer Test, der viele Anwendungen findet, wurde von WILCOXON [1] sowie von MANN und WHITNEY [1] konstruiert.

Es seien X_1 und X_2 unabhängige Variable, und es sei die Verteilungsfunktion $F_i(x)$ der Zufallsvariablen $X_i\ (i = 1, 2)$ stetig. Zunächst wollen wir annehmen, daß für jedes reelle x die Gleichung $F_1(x) = F_2(x)$ gilt. Ferner mögen die x_{1l} $(l = 1, 2, \ldots, n_1)$ und $x_{2j}\ (j = 1, 2, \ldots, n_2)$ sowie die $y_m\ (m = 1, 2, \ldots, n_1 + n_2)$ die gleiche Bedeutung haben wie in 12.6.A.

Wir wollen berechnen, wieviel Nullen der Null-Eins-Folge $\{y_m\}$ auf jede Eins folgen, und dann die Summe dieser Anzahlen bestimmen. Diese Summe bezeichnen wir mit u, d. h., u ist die Anzahl der Paare (x_{1l}, x_{2j}), für die die Ungleichung $x_{2j} \leqq x_{1l}$ gilt. Das ist der beobachtete Wert der *Mann-Whitneyschen Stichprobenfunktion* U.

Die Bestimmung der exakten Verteilung der Stichprobenfunktion U ist äußerst mühevoll. Gestützt auf die Rekursionsformel, die wir nachstehend herleiten wollen, haben MANN und WHITNEY [1] eine Tafel der Verteilung der Stichprobenfunktion U für den Fall $n_1 \leqq n_2 \leqq 8$ erarbeitet.

Wir bezeichnen mit $\overline{p}_{n_1 n_2}(u)$ bzw. $p_{n_1 n_2}(u)$ die Anzahl solcher Folgen aus n_1 Nullen und n_2 Einsen, in denen jede Eins u Nullen vorausgeht, bzw. die Wahrscheinlichkeit für das Auftreten einer solchen Folge. Die Gesamtheit aller Folgen aus n_1 Nullen und n_2 Einsen beträgt $\binom{n_1 + n_2}{n_1}$, und jede Folge tritt mit der gleichen Wahrscheinlichkeit auf. Es gilt daher die Beziehung

$$p_{n_1 n_2}(u) = \frac{n_1!\, n_2!}{(n_1 + n_2)!}\, \overline{p}_{n_1 n_2}(u). \tag{12.6.2}$$

Wir stellen fest, daß $\overline{p}_{n_1 n_2}(u)$ der rekursiven Beziehung

$$\overline{p}_{n_1 n_2}(u) = \overline{p}_{n_1 - 1, n_2}(u - n_2) + \overline{p}_{n_1, n_2 - 1}(u) \tag{12.6.3}$$

genügt, wobei

$$\overline{p}_{n_1 n_2}(u) = 0 \quad \text{für} \quad u < 0 \quad \text{oder} \quad u > n_1 n_2$$

gilt. Die Gleichung (12.6.3) läßt sich wie folgt verifizieren: Wir unterteilen die betrachteten Folgen in zwei Gruppen. Die erste Gruppe besteht aus den Folgen, bei denen rechts an der letzten Stelle eine Null steht, die zweite aus denen, bei denen an der gleichen Stelle eine Eins steht.

Wir bezeichnen mit $A_{n_1 n_2}(u)$ bzw. $B_{n_1 n_2}(u)$ die Anzahl der Folgen der ersten bzw. der zweiten Gruppe, in denen die Einsen u Nullen vorausgehen. Wir haben dann

$$\overline{p}_{n_1 n_2}(u) + A_{n_1 n_2}(u) + B_{n_1 n_2}(u). \tag{12.6.4}$$

Nun tilgen wir in den Folgen der ersten Gruppe die letzte Null. Diese Folgen enthalten jetzt je $n_1 - 1$ Nullen und je n_2 Einsen, und jede Eins geht nunmehr einer Null weniger voraus als in der ursprünglichen Folge dieser Gruppe. Unter Beachtung der Tatsache, daß n_2 Einsen vorhanden sind, erhalten wir somit

$$A_{n_1 n_2}(u) = \overline{p}_{n_1 - 1, n_2}(u - n_2). \tag{12.6.5}$$

Nun tilgen wir in den Folgen der zweiten Gruppe die letzten Einsen. Diese Folgen haben dann jeweils n_1 Nullen und $n_2 - 1$ Einsen. Da die Tilgung der Einsen an der letzten Stelle rechts den Wert u nicht beeinflußt, haben wir

$$B_{n_1 n_2}(u) = \bar{p}_{n_1, n_2-1}(u). \tag{12.6.6}$$

Aus (12.6.4) bis (12.6.6) erhalten wir nun die Beziehung (12.6.3).

Sodann leiten wir unter Anwendung einer von VAN DANTZIG [1] eingeführten Methode die Formeln für $E(U)$ und $D^2(U)$ her, wobei wir nicht voraussetzen, daß $F_1(x) \equiv F_2(x)$ ist.

Wir setzen

$$p = P(X_2 - X_1 < 0) = P(X_2 - X_1 \leqq 0). \tag{12.6.7}$$

Aus (6.5.7) folgt

$$p = \int_{-\infty}^{\infty} F_2(x)\, dF_1(x). \tag{12.6.8}$$

Wir definieren die Funktionen $g(z)$ und ξ_{lj} auf folgende Weise:

$$g(z) = \begin{cases} 1 & \text{für} \quad z > 0, \\ 0 & \text{für} \quad z \leqq 0, \end{cases} \tag{12.6.9}$$

$$\xi_{lj} = g(x_{1l} - x_{2j}) \qquad (l = 1, \ldots, n_1; \quad j = 1, \ldots, n_2). \tag{12.6.9'}$$

Es ist somit $\xi_{lj} = 1$, wobei $x_{1l} > x_{2j}$ ist; hieraus folgt, daß sich die Mann-Whitneysche Stichprobenfunktion U in der Form

$$U = \sum_{l=1}^{n_1} \sum_{j=1}^{n_2} \xi_{lj} \tag{12.6.10}$$

schreiben läßt. Aus (12.6.7) und (12.6.9') erhalten wir

$$E(\xi_{lj}) = E(\xi_{lj}^2) = p. \tag{12.6.11}$$

Da ξ_{lj} und $\xi_{l_1 j_1}$ für $l \neq l_1$ und $j \neq j_1$ unabhängig sind, ergibt sich

$$E(\xi_{lj}\xi_{l_1 j_1}) = p^2 \quad (l \neq l_1, j \neq j_1). \tag{12.6.12}$$

Nunmehr setzen wir

$$\gamma^2 = \int_{-\infty}^{\infty} \big(F_2(x) - p\big)^2\, dF_1(x), \tag{12.6.13}$$

$$\varphi^2 = \int_{-\infty}^{\infty} \big(F_1(x) - (1 - p)\big)^2\, dF_2(x). \tag{12.6.14}$$

Nun sei $j \neq j_1$. Das Produkt $\xi_{lj}\xi_{lj_1}$ ist dann und nur dann gleich 1, wenn die Konjunktion $(x_{2j} - x_{1l} < 0$ und $x_{2j_1} - x_{1l} < 0)$ erfüllt ist, während im entgegengesetzten Fall $\xi_{lj}\xi_{lj_1} = 0$ ist. Wir ermitteln die Wahrscheinlichkeit für die Gültigkeit dieser Konjunktion. Zunächst gilt

$$P(X_{2j} < X_{1l}, X_{2j_1} < X_{1l}) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{x} dF_2(t) \right)^2 dF_1(x) = \int_{-\infty}^{\infty} \big(F_2(x)\big)^2 \, dF_1(x).$$

Wie man aus (12.6.8) und (12.6.13) erkennt, ist das Integral auf der rechten Seite der letzten Gleichung gleich $\gamma^2 + p^2$, woraus

$$E(\xi_{lj}\xi_{lj_1}) = \gamma^2 + p^2 \qquad (j \neq j_1) \tag{12.6.15}$$

folgt. Analog erhalten wir

$$E(\xi_{lj}\xi_{l_1j}) = \varphi^2 + p^2 \qquad (l \neq l_1). \tag{12.6.16}$$

Aus den Beziehungen (12.6.10) bis (12.6.12) sowie aus (12.6.15) und (12.6.16) folgt

$$E(U) = n_1 n_2 p, \tag{12.6.17}$$

$$E(U^2) = \sum_{l=1}^{n_1} \sum_{l_1=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j_1=1}^{n_2} E(\xi_{lj}\xi_{l_1j_1}) = n_1 n_2 \big(p + (n_2 - 1)(\gamma^2 + p^2)$$
$$+ (n_1 - 1)(\varphi^2 + p^2) + (n_1 - 1)(n_2 - 1)p^2\big).$$

Hieraus und aus (12.6.11) erhalten wir

$$D^2(U) = n_1 n_2 \big((n_1 - 1)\varphi^2 + (n_2 - 1)\gamma^2 + p(1 - p)\big). \tag{12.6.18}$$

Für $F_1(x) \equiv F_2(x)$ erhalten wir durch partielle Integration in (12.6.8), (12.6.13) und (12.6.14)

$$p = \frac{1}{2}, \qquad \gamma^2 = \varphi^2 = \frac{1}{12}.$$

Aus (12.6.17) und (12.6.18) ergibt sich dann

$$E(U) = \frac{1}{2} n_1 n_2, \qquad D^2(U) = \frac{1}{12} n_1 n_2 (n_1 + n_2 + 1). \tag{12.6.19}$$

Nun betrachten wir die Stichprobenfunktion

$$V = \frac{U - E(U)}{\sqrt{D^2(U)}}. \tag{12.6.20}$$

MANN und WHITNEY [1] haben für den Spezialfall $F_1(x) \equiv F_2(x)$ bewiesen, daß für $n_1 \to \infty$ und $n_2 \to \infty$ die Stichprobenfunktion V die asymptotische Normalverteilung $N(0;1)$ hat. LEHMANN [2] bewies das gleiche für den allgemeinen Fall, d. h. für $F_1(x) \not\equiv F_2(x)$, jedoch nur dann, wenn $0 < p < 1$ sowie $\frac{n_2}{n_1} \to s$ $(s > 0)$ und $n_1 \to \infty$ ist.

Die Tabellen von MANN und WHITNEY [1] zeigen, daß für $F_1(x) \equiv F_2(x)$ die Approximation bereits für $n_1 = n_2 = 8$ ausreichend genau ist.

Wir wollen nun einen Test behandeln, der auf der Verteilung der Stichprobenfunktion U beruht. Zu diesem Zweck führen wir die folgende Definition ein:

Definition 12.6.1. Die Zufallsvariable X_i $(i = 1, 2)$ möge die Verteilungsfunktion $F_i(x)$ haben. Wir sagen, daß die Zufallsvariable X_2 *stochastisch größer als* X_1 ist, wenn für jedes reelle x, für das $0 < F_1(x) < 1$ gilt, die Ungleichung

$$F_1(x) \geqq F_2(x) \tag{12.6.21}$$

erfüllt ist.

Wir wollen nun annehmen, daß wir zwei Zufallsvariable X_1 und X_2 vorliegen haben und feststellen wollen, ob X_2 stochastisch größer ist als X_1. Wir setzen die Hypothese H_0 voraus, derzufolge die Gleichung $F_1(x) = F_2(x)$ für jedes reelle x erfüllt ist. Diese Hypothese verifizieren wir mit Hilfe der Mann-Whitneyschen Stichprobenfunktion U auf folgende Weise: Wir greifen aus der i-ten Gesamtheit $(i = 1, 2)$ eine n_i-gliedrige Stichprobe heraus. Ferner nehmen wir an, daß n_1 und n_2 nicht größer sind als 8. Wir bestimmen aus der Tabelle von MANN und WHITNEY [1] den Wert u_α derart, daß bei Gültigkeit der Hypothese H_0 die Beziehung $P(U \leqq u_\alpha) \leqq \alpha$ erfüllt ist, wobei α die angenommene Sicherheitswahrscheinlichkeit ist. Sodann bestimmen wir aus den gewonnenen Stichproben den Wert u der Stichprobenfunktion U. Ist hierbei $u \leqq u_\alpha$, so lassen wir die Hypothese H_0 fallen, denn die Tatsache, daß die Anzahl der beobachteten Paare (x_{1l}, x_{2j}), für die $x_{2j} < x_{1l}$ gilt, klein ist, deutet anscheinend darauf hin, daß in der Tat die Zufallsvariable X_2 stochastisch größer ist als X_1.

Wenn n_1 und n_2 größer sind als 8, verwenden wir die Normalapproximation, d. h., wir berechnen den beobachteten Wert v der Stichprobenfunktion V, die durch (12.6.20) gegeben ist, wobei wir $E(U)$ und $D^2(U)$ aus den Formeln (12.6.19) bestimmen, und lassen die Hypothese H_0 fallen, wenn $v \leqq v_\alpha$ ist; hierbei ist $\Phi(v_\alpha) = \alpha$ und α die Sicherheitswahrscheinlichkeit.

Beispiel 12.6.2. Die Tabelle 12.6.1 gibt in acres (1 acre = 0,40468 ha) für 34 Dörfer des Gebietes von Laknau (Lucknow) in Indien die Flächen an, die in den Sommern 1936 und 1937 mit Weizen bebaut wurden. Die Hypothese, die wir testen wollen, möge darin bestehen, daß die Variable „Weizenfläche" in den Sommern 1936 und 1937 die gleiche Verteilung hat. Wir wenden den Test von MANN-WHITNEY zur Verifizierung dieser Hypothese an. Aus der Tabelle 12.6.1 erhalten wir, nach Ordnung der Zahlen x_{1l} und x_{2j}, wenn wir den Beobachtungen aus dem Jahre 1936 die Zahl Null und den Beobachtungen aus dem Jahre

1937 die Zahl Eins zuordnen, die Ziffernfolge (Null-Eins-Folge)

$\{y_m\} = 0110110100000111011011011100001001110100110001000111101101100011 0010.$

Hieraus erhalten wir $u = 596$ und $v = \dfrac{596 - 17 \cdot 34}{\sqrt{\dfrac{34 \cdot 34 \cdot 69}{12}}} = \dfrac{18}{17\sqrt{23}} \approx 0{,}221.$

Wegen $\Phi(0{,}221) \approx 0{,}5875$ liegt kein Grund vor, die getestete Hypothese zu verwerfen. Wir weisen darauf hin, daß in der Tabelle 12.6.1 manche Zahlen mehrfach auftreten. Das liegt daran, daß die beobachteten Werte mit einer Genauigkeit bis zu einem halben acre angegeben wurden.

Tabelle 12.6.1.

Nummer des Dorfes	1936 x_{1l}	1937 x_{2j}	Nummer des Dorfes	1936 x_{1l}	1937 x_{2j}
1	75	52	18	45	27
2	163	149	19	564	515
3	326	289	20	238	249
4	442	381	21	92	85
5	254	278	22	247	221
6	125	111	23	134	133
7	559	634	24	131	144
8	254	278	25	129	103
9	101	112	26	192	179
10	359	355	27	663	330
11	109	99	28	236	219
12	481	498	29	73	62
13	125	111	30	62	79
14	5	6	31	71	60
15	427	399	32	137	100
16	78	79	33	196	141
17	78	105	34	255	265

Die Zahlenwerte für dieses Beispiel sind dem Buch von SUKHATME [3] entnommen. Auf den Test von MANN-WHITNEY werden wir noch einmal im Kapitel 16 zurückkommen.

WILCOXON [1] betrachtete für $n_1 = n_2$, früher als MANN und WHITNEY, eine Stichprobenfunktion, die mit der Mann-Whitneyschen Stichprobenfunktion U linear verknüpft ist, und zwar die Stichprobenfunktion $S = \dfrac{1}{2}\, n_2(2n_1 + n_2 + 1) - U$. Aus diesem Grunde findet man in der Literatur für den betrachteten Test auch häufig die Bezeichnung *Test von* WILCOXON-MANN-WHITNEY. Wir weisen darauf hin, daß $S = m_1 + m_2 + \cdots + m_{n_2}$ gilt, wobei die m_r $(r = 1, 2, \ldots, n_2)$ die Ränge derjenigen y_m sind, die den Elementen x_{2j} entsprechen.

Die Sicherheitswahrscheinlichkeiten für den Test von WILCOXON-MANN-WHITNEY wurden von FIX und HODGES [1] angegeben.

Die Verallgemeinerung dieses Tests auf den Fall von k Stichproben ($k > 2$) gehen auf KRUSKAL [1], RIJKOORT [1] und PFANZAGL [1] zurück.

C. Die Theorie der nichtparametrischen Tests hat sich in den letzten Jahren sehr stark entwickelt. FRASER [1] widmete dieser Theorie eine umfangreiche Monographie. Ferner seien hier auch die Arbeiten von WOLFOWITZ [4], GICHMAN, GNEDENKO und SMIRNOW [1] sowie SCHMETTERER [3] erwähnt, die viele wichtige Informationen über neue Untersuchungen auf diesem Gebiet enthalten.

Neben den rein parametrischen und den rein nichtparametrischen Tests werden in der Literatur Tests gemischter Natur betrachtet, und zwar Tests, bei denen die Funktionsform der Verteilungsfunktion als unbekannt angesehen wird, doch ist der Gegenstand der zu testenden Hypothese nicht die Funktionsform der Verteilungsfunktion, sondern der Vergleich der Werte der unbekannten Parameter mehrerer Grundgesamtheiten. Solche Tests, die die Erwartungswerte und Varianzen mehrerer Grundgesamtheiten mit unbekannten Verteilungsfunktionen betreffen, wurden von MOSTELLER [1], SADOWSKI [1], ROSENBAUM [1] sowie LUKASZEWICZ und SADOWSKI [1] angegeben. (PAULSON [1] betrachtete ähnliche Probleme für mehrere normale Gesamtheiten.)

12.7. Unabhängigkeitstests in Kontingenztafeln

Wir stellen uns vor, daß die Elemente einer Stichprobe im Hinblick auf zwei Merkmale X und Y klassifiziert wurden. Dabei wurde der Bereich der möglichen Werte der Merkmale in r bzw. s Gruppen aufgeteilt. Wir bezeichnen mit n den Umfang der Stichprobe, mit n_{ik} die Anzahl der Elemente, welche zur i-ten ($i = 1, 2, \ldots, r$) Gruppe hinsichtlich des Merkmals X und zur k-ten ($k = 1, 2, \ldots, s$) Gruppe hinsichtlich des Merkmals Y gehören. Weiter setzen wir

$$n_{i\cdot} = \sum_{k=1}^{s} n_{ik}, \tag{12.7.1}$$

$$n_{\cdot k} = \sum_{i=1}^{r} n_{ik}. \tag{12.7.2}$$

Selbstverständlich gilt die Gleichung

$$n = \sum_{i=1}^{r} \sum_{k=1}^{s} n_{ik}.$$

Die so klassifizierten Stichprobenelemente stellen wir in Tabelle 12.7.1 dar (*Kontingenztafel*).

Tabelle 12.7.1.

i \ k	1 2 … s	zusammen $n_{i\cdot}$
1	n_{11} n_{12} … n_{1s}	$n_{1\cdot}$
2	n_{21} n_{22} … n_{2s}	$n_{2\cdot}$
. . . .	. . . .	. . . .
r	n_{r1} n_{r2} … n_{rs}	$n_{r\cdot}$
zusammen $n_{\cdot k}$	$n_{\cdot 1}$ $n_{\cdot 2}$ … $n_{\cdot s}$	n

Wir stellen die Hypothese H_0 auf, daß in der Gesamtheit, aus der die betrachtete Stichprobe stammt, die Merkmale X und Y unabhängig sind. Wenn wir mit p_{ik} die Wahrscheinlichkeit dafür bezeichnen, daß ein aufs Geratewohl aus der Gesamtheit herausgegriffenes Element zur i-ten Gruppe hinsichtlich des Merkmals X und zur k-ten Gruppe hinsichtlich des Merkmals Y gehört, und mit $p_{i\cdot}$ und $p_{\cdot k}$ die entsprechenden Randwahrscheinlichkeiten, dann ist nach Hypothese H_0 für jedes Paar (i, k)

$$p_{ik} = p_{i\cdot}\, p_{\cdot k}, \tag{12.7.3}$$

wobei

$$\sum_{i=1}^{r} p_{i\cdot} = \sum_{k=1}^{s} p_{\cdot k} = 1 \tag{12.7.4}$$

gilt. Die so formulierte Hypothese H_0 präzisiert nicht die Werte der $r + s$ unbekannten Parameter, d. h. die Werte $p_{\cdot k}$ und $p_{i\cdot}$. Da man zwei von ihnen aus den Formeln (12.7.4) bestimmen kann, bleiben noch $r + s - 2$ Parameter unbekannt. Um zur Verifikation der Hypothese H_0 einen χ^2-Test benutzen zu können, bestimmen wir die unbekannten Parameter nach dem Maximum-Likelihood-Prinzip aus der Stichprobe. Es gilt

$$L = \prod_{i=1}^{r} \prod_{k=1}^{s} p_{ik}^{n_{ik}} = \prod_{i=1}^{r} \prod_{k=1}^{s} (p_{i\cdot}\, p_{\cdot k})^{n_{ik}} = \prod_{i=1}^{r} \prod_{k=1}^{s} p_{i\cdot}^{n_{ik}}\, p_{\cdot k}^{n_{ik}}.$$

Unter Berücksichtigung der Bezeichnungen (12.7.1) und (12.7.2) erhalten wir

$$L = \prod_{i=1}^{r} p_{i\cdot}^{n_{i\cdot}} \prod_{k=1}^{s} p_{\cdot k}^{n_{\cdot k}}.$$

Aus den Gleichungen (12.7.4) bestimmen wir $p_{r\cdot}$ und $p_{\cdot s}$:

$$p_{r\cdot} = 1 - \sum_{i=1}^{r-1} p_{i\cdot},$$

$$p_{\cdot s} = 1 - \sum_{k=1}^{s-1} p_{\cdot k}.$$

Daraus erhalten wir

$$L = \left(1 - \sum_{i=1}^{r-1} p_{i\cdot}\right)^{n_{r\cdot}} \left(1 - \sum_{k=1}^{s-1} p_{\cdot k}\right)^{n_{\cdot s}} \prod_{i=1}^{r-1} p_{i\cdot}^{n_{i\cdot}} \prod_{k=1}^{s-1} p_{\cdot k}^{n_{\cdot k}}$$

und

$$\log L = n_{r\cdot} \log\left(1 - \sum_{i=1}^{r-1} p_{i\cdot}\right) + n_{\cdot s} \log\left(1 - \sum_{k=1}^{s-1} p_{\cdot k}\right) + \sum_{i=1}^{r-1} n_{i\cdot} \log p_{i\cdot} + \sum_{k=1}^{s-1} n_{\cdot k} \log p_{\cdot k}.$$

Wir haben somit das Gleichungssystem

$$\frac{\partial \log L}{\partial p_{i\cdot}} = -\frac{n_{r\cdot}}{1 - \sum\limits_{i=1}^{r-1} p_{i\cdot}} + \frac{n_{i\cdot}}{p_{i\cdot}} = \frac{n_{i\cdot}}{p_{i\cdot}} - \frac{n_{r\cdot}}{p_{r\cdot}} = 0 \qquad (i = 1, 2, \ldots, r-1),$$

$$\frac{\partial \log L}{\partial p_{\cdot k}} = -\frac{n_{\cdot s}}{1 - \sum\limits_{k=1}^{s-1} p_{\cdot k}} + \frac{n_{\cdot k}}{p_{\cdot k}} = \frac{n_{\cdot k}}{p_{\cdot k}} - \frac{n_{\cdot s}}{p_{\cdot s}} = 0 \qquad (k = 1, 2, \ldots, s-1).$$

Setzen wir

$$A = \frac{n_{r\cdot}}{p_{r\cdot}}, \qquad B = \frac{n_{\cdot s}}{p_{\cdot s}},$$

so ergibt sich

$$p_{i\cdot} = \frac{n_{i\cdot}}{A} \qquad (i = 1, 2, \ldots, r),$$

$$p_{\cdot k} = \frac{n_{\cdot k}}{B} \qquad (k = 1, 2, \ldots, s).$$

Aus der Gleichung (12.7.4) erhalten wir

$$\sum_{i=1}^{r} p_{i\cdot} = \frac{\sum\limits_{i=1}^{r} n_{i\cdot}}{A} = \frac{n}{A} = 1,$$

$$\sum_{k=1}^{s} p_{\cdot s} = \frac{\sum\limits_{k=1}^{s} n_{\cdot k}}{B} = \frac{n}{B} = 1.$$

also ist $A = B = n$. Endgültig ergibt sich

$$p_{i\cdot} = \frac{n_{i\cdot}}{n} \qquad (i = 1, 2, \ldots, r),$$

$$p_{\cdot k} = \frac{n_{\cdot k}}{n} \qquad (k = 1, 2, \ldots, s). \tag{12.7.5}$$

Wenn wir (12.4.2.), (12.7.3) und (12.7.5) berücksichtigen, erhalten wir

$$\chi^2 = n \sum_{i=1}^{r} \sum_{k=1}^{s} \frac{\left(n_{ik} - \frac{n_{i\cdot} n_{\cdot k}}{n}\right)^2}{n_{i\cdot} n_{\cdot k}}. \tag{12.7.6}$$

Da wir aus der Stichprobe $r + s - 2$ Parameter bestimmt haben, ergibt sich nach dem Fisherschen Satz (siehe 12.4), daß die durch (12.7.6) bestimmte Zufallsvariable χ^2 in Annäherung eine χ^2-Verteilung mit

$$rs - (r + s - 2) - 1 = rs - r - s + 1 = (r - 1)(s - 1)$$

Freiheitsgraden hat. Folglich können wir einen χ^2-Test zum Nachprüfen der Hypothese H_0 benutzen, die die Unabhängigkeit der Variablen X und Y behauptet.

Wenn $r = s = 2$ ist, nennen wir die Kontingenztafel Tabelle 12.7.1 *vierteilig*. Die Formel für χ^2 hat hier die Gestalt

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\cdot} n_{2\cdot} n_{\cdot 1} n_{\cdot 2}}. \tag{12.7.7}$$

Beispiel 12.7.1. Wir stellen die Hypothese H_0 auf, daß bei Männern die Augenfarbe X und die Haarfarbe Y voneinander unabhängige Merkmale sind. Diese Hypothese wollen wir nachprüfen, indem wir die Augen- und Haarfarbe an 6800 Männern untersuchen und die

Tabelle 12.7.2.

k (Haar) / i (Augen)	1 (hellblond)	2 (dunkelblond)	3 (schwarz)	4 (rot)	zusammen $n_{i\cdot}$
1 (blau)	1768	807	189	47	2811
2 (grau oder grün)	946	1387	746	53	3132
3 (braun)	115	438	288	16	857
zusammen $n_{\cdot k}$	2829	2632	1223	116	6800

Angaben in Tabelle 12.7.2 zusammenstellen, die aus M. G. KENDALL [1] entnommen wurde. Die Stichprobe teilen wir hinsichtlich der Augenfarbe in drei und hinsichtlich der Haarfarbe in vier Gruppen ein. Aus Tabelle 12.7.2 berechnen wir

$$\chi^2 = 6800 \sum_{i=1}^{3} \sum_{k=1}^{4} \frac{\left(n_{ik} - \dfrac{n_{i\cdot}\, n_{\cdot k}}{6800}\right)}{n_{i\cdot}\, n_{\cdot k}} = 1075{,}2.$$

Die Anzahl der Freiheitsgrade beträgt $(r - 1)(s - 1) = 2 \cdot 3 = 6$. In den Tafeln der χ^2-Verteilung finden wir, daß für sechs Freiheitsgrade $P(\chi^2 \geqq 1075{,}2) < 0{,}000001$ ist. Wir lehnen deshalb die Hypothese H_0 ab.

12.8. Aufgaben und Ergänzungen

1. In zwei einfachen Stichproben aus zwei normalen Gesamtheiten mit den Verteilungen $N(0; \sigma_1)$ und $N(0; \sigma_2)$, wobei σ_1 und σ_2 unbekannt sind, wurde $s_1 = 1$ und $s_2 = 3$ beobachtet. Die Stichprobenumfänge waren $n_1 = 9$ bzw. $n_2 = 7$. Mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}02$ teste man die Hypothesen

   a) $H_0(\sigma_1 = \sigma_2 = 1)$, b) $H_0(\sigma_1 = 1, \sigma_2 = 2)$ und c) $H_0\left(\sigma_1 = \dfrac{1}{2}, \sigma_2 = 1\right)$.

   Hinweis. Man verwende die Aufgabe 9.12.7.

2. In einer siebengliedrigen einfachen Stichprobe aus einer normalen dreidimensionalen Grundgesamtheit wurden die folgenden Werte für die Stichprobenkorrelationskoeffizienten ermittelt: $r_{12} = 0{,}2$, $r_{13} = 0{,}4$, $r_{23} = -0{,}35$. Man teste für die Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ die Hypothese $H_0(\varrho_{12} = \varrho_{13} = \varrho_{23} = 0)$.

   Hinweis. Man verwende die Aufgabe 9.12.10.

3. In einer einfachen sechsgliedrigen Stichprobe aus einer dreidimensionalen normalen Gesamtheit hat sich der partielle Korrelationskoeffizient $r_{12\cdot 3} = 0{,}26$ ergeben. Mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ teste man die Hypothese, daß der partielle Korrelationskoeffizient $\varrho_{12\cdot 3}$ der Grundgesamtheit gleich Null ist.

   Hinweis. Man verwende die Aufgabe 9.12.12.

4. Bei einer einfachen zehngliedrigen Stichprobe aus einer vierdimensionalen normalen Gesamtheit ergab sich $r_{1(234)} = 0{,}3$. Widerspricht dieses Ergebnis bei der Sicherheitswahrscheinlichkeit $\alpha = 0{,}01$ der Hypothese, daß $\varrho_{12} = \varrho_{13} = \varrho_{14} = \varrho_{23} = \varrho_{24} = 0$ ist?

   Hinweis. Man verwende die Aufgabe 9.12.13.

5. Bei der Beobachtung einer dünnen Schicht einer Goldlösung registrierte SVEDBERG in gleichen Zeitintervallen die Anzahl der Goldpartikeln im Mikroskopfeld. In folgender Tabelle bezeichnet n_j die Anzahl der Zeitintervalle, in denen SVEDBERG j Goldpartikeln registrierte. Man fasse n_j als Werte der Zufallsvariablen N auf und überprüfe mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}01$ die Hypothese, daß N die Poissonsche Verteilung mit dem Parameter λ hat, wobei λ gleich der beobachteten mittleren Anzahl der Partikeln in einem Zeitintervall ist.

j	0	2	3	4	5	6	7	8	insgesamt
n_j	112	168	130	68	32	5	1	1	517

6. Man zeige, daß die durch die Formel (12.4.2) definierte Stichprobenfunktion χ^2 die Momente

$$E(\chi^2) = r - 1, \qquad D^2(\chi^2) = 2(r-1) + \frac{A - (r^2 + 2r - 2)}{n}$$

hat, wobei $A = \sum_{k=1}^{r} \frac{1}{\pi_k}$ ist. Insbesondere gilt für $\pi_k = \frac{1}{r}$ $(k = 1, 2, \ldots, r)$

$$D^2(\chi^2) = \frac{2(r-1)(n-1)}{n}.$$

7. In der Formel (12.4.2) möge r für $n \to \infty$ über alle Grenzen wachsen. Man beweise: Gilt hierbei

$$\lim_{n\to\infty} \min_{1 \leq k \leq r} n\pi_k = \infty,$$

so besitzt die Stichprobenfunktion $(\chi^2 - r)/\sqrt{2r}$ die asymptotische Normalverteilung $N(0; 1)$ (Tumanjan [1], Verallgemeinerung durch Gichman [4]).

8. An zehn Männern wurden Messungen des Kniesehnenreflexes unter zwei verschiedenen Bedingungen I und II vorgenommen. Im ersten Fall (I) hatte die Versuchsperson unmittelbar vor dem Schlag mit dem Reflexhammer ein Handdynamometer zusammenzupressen. Im zweiten Fall (II) wurde der Kniesehnenreflex im entspannten Zustand erregt. Die Größe der Reflexbewegung in Grad gemessen gibt die nachstehende Tabelle an (nach Guilford [1]):

I	19	19	26	15	18	30	18	30	26	28
II	14	19	30	7	13	20	17	29	18	21

Unter Verwendung des Tests von Mann-Whitney mit der Sicherheitswahrscheinlichkeit $\alpha = 0{,}05$ überprüfe man, ob der Knieausschlag unter den Bedingungen I und II die gleiche Verteilung hat.

9. Jedes Element einer Grundgesamtheit vom Umfang N kann a) die Eigenschaft A und die Eigenschaft B haben, b) die Eigenschaft A besitzen, die Eigenschaft B hingegen nicht, c) die Eigenschaft A nicht aufweisen, dagegen jedoch die Eigenschaft B, d) weder die Eigenschaft A noch die Eigenschaft B besitzen. Die entsprechenden Elementanzahlen in den einzelnen Gruppen seien N_{11}, N_{12}, N_{21} und N_{22}. Der Ausdruck

$$q = \frac{N_{11}N_{22} - N_{12}N_{21}}{N_{11}N_{22} + N_{12}N_{21}}$$

heißt der *Assoziationskoeffizient* (Yule [1], [2]).
Wir wollen das Aufweisen der Eigenschaft A als zufälliges Ereignis A und das Aufweisen der Eigenschaft B als zufälliges Ereignis B deuten. Man beweise, daß q bei Unabhängigkeit der zufälligen Ereignisse A und B gleich 0 ist. Wie sind die Fälle $q = 1$ und $q = -1$ zu deuten?

10. (Fortsetzung). In n-gliedrigen einfachen Stichproben aus der in Aufgabe 9 betrachteten Gesamtheit erhalten wir n_{11}, n_{12}, n_{21} bzw. n_{22} Elemente der einzelnen Kategorien. Der Assoziationskoeffizient der Stichprobe hat die Form

$$S = \frac{n_{11}\,n_{22} - n_{12}\,n_{21}}{n_{11}\,n_{22} + n_{12}\,n_{21}}.$$

Man zeige, daß dann

$$D^2(S) = \frac{(1-q^2)^2}{4}\left(\frac{1}{N_{11}} + \frac{1}{N_{12}} + \frac{1}{N_{21}} + \frac{1}{N_{22}}\right)$$

gilt.

11. Man bestimme den Assoziationskoeffizienten für eine Stichprobe, die den Zusammenhang zwischen der Choleraimpfung und der Morbidität ermitteln sollte, aus der nachstehenden Tabelle (Greenwood und Yule [1]):

	nicht erkrankt	erkrankt	insgesamt
geimpft	276	3	279
nicht geimpft	473	66	539
insgesamt	749	69	818

Man berechne $D^2(S)$ durch Einsetzen der beobachteten Anzahlen anstelle der theoretischen.

13. THEORIE DER SCHÄTZFUNKTIONEN

13.1. Einleitende Bemerkungen

In 12.4 und 12.7 waren wir schon auf die Frage nach der Abschätzung unbekannter Parameter einer Gesamtheit auf Grund von Stichprobenergebnissen gestoßen und hatten eine Methode kennengelernt, die diese Abschätzungen gewährleistet. Die allgemeine Theorie, die sich mit dieser Problemstellung befaßt, nennt man *Schätztheorie.*

Um einen unbekannten Parameter Q auf Grund von Stichprobenergebnissen abschätzen zu können, geht man folgendermaßen vor. Man berechnet aus der Stichprobe den Wert u einer gewissen Stichprobenfunktion U, deren Verteilung von diesem Parameter Q abhängt, und nimmt diesen Wert u als Abschätzung des unbekannten Parameters Q. Wir sprechen dann von einer *Punktschätzung.* Die Stichprobenfunktion U sowie ihren beobachteten Wert u bezeichnen wir als *Schätzfunktion des unbekannten Parameters.* Gewöhnlich geht es aus dem Text hervor, ob der Terminus Schätzfunktion sich auf U oder auf u bezieht.

Man kann desgleichen nach gewissen Intervallen suchen, die den unbekannten Parameter enthalten. In diesem Fall sprechen wir von einer *Intervallschätzung.*

Wir beginnen die speziellen Betrachtungen mit einer Klassifikation der Schätzfunktionen.

13.2. Konsistente Schätzfunktionen

Mit $F(x, Q)$ sei die Verteilungsfunktion des Merkmals X der Gesamtheit bezeichnet; der unbekannte Parameter Q ist zu schätzen.

Wir betrachten eine Schätzfunktion U_n des Parameters Q, wobei mit n der Stichprobenumfang bezeichnet ist.

Offenbar ist oft der auf Grund einer Stichprobe geschätzte Wert des unbekannten Parameters vom wirklichen Wert dieses Parameters verschieden. Es ist jedoch natürlich zu verlangen, daß mit Anwachsen des Stichprobenumfangs die Wahrscheinlichkeit dafür, daß der geschätzte Wert nahe dem wirklichen Parameterwert liegt, gegen Eins strebt.

Definition 13.2.1. Wir sagen, daß die Folge $\{U_n\}$ $(n = 1, 2, \ldots)$ von Schätzfunktionen des Parameters Q *konsistent* ist, wenn sie für $n \to \infty$ stochastisch gegen Q konvergiert.

Ist U_n Glied einer konsistenten Folge $\{U_n\}$ von Schätzfunktionen des Parameters Q, so wollen wir kurz sagen, daß U_n eine *konsistente Schätzfunktion des Parameters* Q ist.

Gemäß der Definition der stochastischen Konvergenz (siehe 6.2) ist eine Schätzfunktion U_n konsistent, wenn für jedes $\varepsilon > 0$ die Beziehung

$$\lim_{n\to\infty} P(|U_n - Q| > \varepsilon) = 0 \tag{13.2.1}$$

besteht.

Den geschätzten Wert des Parameters Q bezeichnen wir häufig mit Q^*.

Wir bringen einige Beispiele für konsistente Schätzfunktionen.

Beispiel 13.2.1. Der Mittelwert einer Gesamtheit, deren Merkmal X normal $N(m; 1)$ verteilt ist, sei unbekannt. Aus einer einfachen Stichprobe vom Umfang n bestimmen wir den Mittelwert x und nehmen $m^* = \bar{x}$ an. Mit anderen Worten, wir wählen als Schätzfunktion die Stichprobenfunktion

$$\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k,$$

wobei die Zufallsvariablen X_k unabhängig sind und dieselbe Verteilung wie die Zufallsvariable X haben. Wie schon bekannt (siehe 9.11), konvergiert $\bar{X}$ stochastisch gegen m, d. h., $\bar{X}$ ist eine konsistente Schätzfunktion.

Beispiel 13.2.2. Eine andere konsistente Schätzfunktion des unbekannten Mittelwerts m einer normalen Gesamtheit in einfachen Stichproben ist die Stichprobenmediane, das heißt der beobachtete Wert y der Positionsstichprobenfunktion $\zeta_k^{(n)}$, wobei $k = \left[\frac{n}{2}\right] + 1$ ist. Die Mediane ist deshalb eine konsistente Schätzfunktion, weil für die Normalverteilung der Zentralwert mit dem Mittelwert zusammenfällt und weil nach Satz 10.4.1 die Stichprobenmediane stochastisch gegen die Gesamtheitsmediane konvergiert.

Aus den obigen Beispielen ersehen wir, daß in einfachen Stichproben die Stichprobenmediane und der Mittelwert konsistente Schätzfunktionen des Parameters m sind. Nun kann man fragen, welche dieser beiden eine bessere Schätzung für den Parameter m liefert, d. h. welche häufiger einen näher an m liegenden Wert m^* ergibt? Auch kann man fragen, welcher dieser Werte in dem Sinne besser ist, daß er Erwartungswerte $E(m^*)$ ergibt, die näher am wirklichen Wert von m liegen? Mit diesen Fragen wollen wir uns jetzt näher befassen.

13.3. Erwartungstreue Schätzfunktionen

Die Konsistenz einer Schätzfunktion drückt deren Grenzeigenschaften aus. Zur Charakterisierung einer Abschätzung kann sie nur benutzt werden, wenn große Stichproben vorliegen. Jetzt wollen wir den Begriff einer erwartungstreuen Schätzfunktionen angeben, der auch Schätzfunktionen in Stichproben eines beliebigen Umfangs (klein oder groß) betrifft.

Ein geschätzter Wert, den man aus einer einzelnen Stichprobe gewonnen hat, kann vom wirklichen Parameterwert verschieden sein. Liegen jedoch Abschätzungen vor, die aus vielen Stichproben stammen, so ist es vernünftig, zu fordern, daß der Erwartungswert dieser Abschätzungen gleich dem wirklichen Wert des unbekannten Parameters ist.

Definition 13.3.1. Es sei D eine Klasse von Verteilungsfunktionen $F(x, Q)$. Die Schätzfunktion U_n des Parameters Q heißt *erwartungstreu in der Klasse D*, wenn für jede Verteilungsfunktion $F(x, Q)$ aus der Klasse D die Relation

$$E(U_n) = Q \qquad (n = 1, 2, \ldots) \tag{13.3.1}$$

besteht. Die Differenz $E(U_n) - Q$ nennen wir die *erwartungsmäßige Abweichung* oder die *Verzerrung der Schätzfunktion*. Man nennt U_n eine *asymptotisch erwartungstreue Schätzfunktion des Parameters Q in der Klasse D*, wenn für jede Verteilungsfunktion $F(x, Q)$ aus der Klasse D die Beziehung

$$\lim_{n\to\infty} E(U_n) = Q \tag{13.3.1'}$$

erfüllt ist.

Beispiel 13.3.1. Wir betrachten eine Gesamtheit, deren Merkmal X beliebig verteilt ist, und wollen den unbekannten Mittelwert $E(X) = m$ abschätzen.

Wir entnehmen dieser Gesamtheit eine Stichprobe vom Umfang n und bestimmen den Wert $\bar{x}$ der Stichprobenfunktion

$$\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k;$$

dabei haben die Variablen X_k dieselbe Verteilung wie die Variable X.

Die Stichprobenfunktion $\bar{X}$ ist eine erwartungstreue Schätzfunktion des Parameters m; wir haben nämlich

$$E(\bar{X}) - E\left(\frac{1}{n} \sum_{k=1}^{n} X_k\right) = \frac{1}{n} \sum_{k=1}^{n} E(X_k) = m. \tag{13.3.2}$$

Somit ist die Stichprobenfunktion $\bar{X}$ eine erwartungstreue Stichprobenfunktion des Erwartungswertes m in der Klasse D aller Verteilungsfunktionen von Zufallsvariablen, deren Momente erster Ordnung existieren.

Die Stichprobenmomente einer Ordnung höher als 1 sind nicht immer erwartungstreue Schätzfunktionen der Gesamtheitsmomente, was wir an einem Beispiel zeigen wollen (siehe auch Aufgabe 9.12.17 bis 9.12.21).

Beispiel 13.3.2. Der unbekannte Parameter einer beliebig verteilten Gesamtheit sei σ^2. Als Schätzfunktion dieses Parameters wählen wir die Stichprobenfunktion

$$S^2 = \frac{1}{n} \sum_{k=1}^{n} X_k^2 - \bar{X}^2. \tag{13.3.3}$$

Dabei sind die Zufallsvariablen X_k unabhängig und haben dieselbe Verteilung wie die Zufallsvariable X. Es gilt

$$E(S^2) = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k^2\right) - E(\overline{X}^2) = E(X^2) - E(\overline{X}^2).$$

Da

$$E(\overline{X}^2) = E\left[\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2\right] = \frac{1}{n^2} E\left(\sum_{k=1}^{n} X_k^2 + \sum_{l,j(l\neq j)} X_l X_j\right)$$

$$= \frac{1}{n} E(X^2) + \frac{n-1}{n}[E(X)]^2$$

ist, erhalten wir also

$$E(S^2) = \frac{n-1}{n} E(X^2) - \frac{n-1}{n}[E(X)]^2 = \frac{n-1}{n}\sigma^2. \tag{13.3.4}$$

Daraus folgt, daß S^2 keine erwartungstreue Schätzfunktion des Parameters σ^2 ist. Wenn wir zum Beispiel für eine Reihe von einfachen Stichproben vom Umfang $n = 2$ den Wert s^2 als Abschätzung von σ^2 wählen, so wird der Durchschnittswert dieser Abschätzungen $\frac{\sigma^2}{2}$ betragen. Jedoch ist S^2 eine asymptotisch erwartungstreue Schätzfunktion des Parameters σ^2.

Wie man leicht bemerkt, stellt die Stichprobenfunktion

$$S_1^2 = \frac{n}{n-1} S^2 \tag{13.3.5}$$

eine erwartungstreue Schätzfunktion des Parameters σ^2 dar. Daraus kann man aber noch nicht schließen, daß für jeden einzelnen Versuch der Wert s_1^2 der Stichprobenfunktion S_1^2 eine genauere Abschätzung des Parameters σ^2 als der Wert s^2 liefert. Die Erwartungstreue ist eine Durchschnittseigenschaft. Wollen wir also in einer langen Reihe von Versuchen durchschnittlich gute Abschätzungen gewinnen, so müssen wir die Formel (13.3.5) anwenden. Dabei ist es jedoch nicht ausgeschlossen, daß die Abschätzung von σ^2 mit Hilfe des Wertes s^2 im Einzelfall näher am wirklichen Wert von σ^2 als s_1^2 liegt. Es besteht also kein Grund, einen der beiden besonders zu bevorzugen.

Offenbar werden sich die Werte von S^2 und S_1^2 nur wenig voneinander unterscheiden, wenn n groß ist.

Wir bemerken, daß bei bekanntem Mittelwert $m = E(X)$ die Schätzfunktion

$$S_0^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2 \tag{13.3.6}$$

des Parameters σ^2 erwartungstreu ist. In der Tat gilt

$$E(S_0^2) = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k^2\right) - m^2 = E(X^2) - m^2 = \sigma^2. \tag{13.3.7}$$

Ist U eine erwartungstreue Schätzfunktion des Parameters Q, so folgt daraus noch nicht, daß jede Funktion $f(U)$ eine erwartungstreue Schätzfunktion des Parameters $f(Q)$ ist. Wir zeigen dies an dem folgenden Beispiel.

Beispiel 13.3.3. Das Merkmal X einer Gesamtheit sei normal $N(m; \sigma)$ verteilt, der Mittelwert m sei bekannt, während die Standardabweichung σ unbekannt sei. Wie wir schon wissen, ist die Stichprobenfunktion S_0^2 eine erwartungstreue Schätzfunktion des Parameters σ^2. Wir zeigen nun, daß die Schätzfunktion S_0 des Parameters σ nicht erwartungstreu ist.

Tatsächlich führen die Formeln (9.4.9) und (5.8.5) zu

$$E(S_0) = \frac{n^{\frac{n}{2}}}{(2\sigma^2)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{-\frac{ns_0^2}{2\sigma^2}} s_0^{2\,\frac{n-1}{2}} \, ds_0^2 = \frac{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\,\Gamma\left(\frac{n}{2}\right)} \sigma \neq \sigma. \tag{13.3.8}$$

Man erkennt leicht, daß die aus einer einfachen Stichprobe berechnete Stichprobenfunktion

$$\frac{\sqrt{n}\,\Gamma\left(\frac{1}{2}\,n\right)}{\sqrt{2}\,\Gamma\left(\frac{1}{2}\,(n+1)\right)} S_0$$

eine erwartungstreue Schätzfunktion des Parameters σ in der Klasse D der Verteilungsfunktionen normal verteilter Zufallsvariabler ist.

B. Der Begriff der erwartungstreuen Schätzfunktion wurde durch Gauss eingeführt (siehe auch David und Neyman [1]). Einerseits braucht eine erwartungstreue Schätzfunktion eines Parameters nicht immer zu existieren, andererseits kann es für ein und denselben Parameter viele erwartungstreue Schätzfunktionen geben. Dem Problem der Existenz und Eindeutigkeit von erwartungstreuen Schätzfunktionen sind unter anderen die Arbeiten von Halmos [1], Lehmann und Scheffé [1], Kolmogoroff [14], Stein [2] und Schmetterer [2] gewidmet.

Im weiteren wollen wir im Buch stets kurz von der *erwartungstreuen Schätzfunktion* sprechen, sofern nicht ausdrücklich die Notwendigkeit besteht, die Klasse D besonders zu kennzeichnen, in der die betreffende Schätzfunktion erwartungstreu ist.

13.4. Erschöpfende Schätzfunktionen

Der Begriff einer erschöpfenden Schätzfunktion, der von Fisher [6] eingeführt wurde, erfordert tiefergehende Betrachtungen.

In 12.4 definierten wir die Likelihood-Funktion L für stetige und für diskrete Zufallsvariable. Wir wollen jetzt eingehender stetige Variable betrachten. Es sei

$f(x, Q)$ die Dichte der Zufallsvariablen X, wobei Q ein unbekannter Parameter sei. X_k $(k = 1, 2, \ldots, n)$ seien unabhängige Zufallsvariable, die ebenso wie X verteilt seien. Ferner seien $x_1, x_2, \ldots, x_n$ die beobachteten Werte der Zufallsvariablen $X_1, X_2, \ldots, X_n$. Also stellt $(x_1, x_2, \ldots, x_n)$ einen Punkt im n-dimensionalen Stichprobenraum dar. Diesen Punkt wollen wir mit A bezeichnen.

Die Likelihood-Funktion L wurde durch die Formel

$$L = f(x_1, Q) f(x_2, Q) \cdots f(x_n, Q) \tag{13.4.1}$$

definiert. Statt der Koordinaten $x_1, x_2, \ldots, x_n$ des Punktes A im Stichprobenraum wollen wir die durch

$$\begin{aligned} u &= \Psi(x_1, x_2, \ldots, x_n), \\ y_j &= \Psi_j(x_1, x_2, \ldots, x_n) \qquad (j = 1, 2, \ldots, n-1) \end{aligned} \tag{13.4.2}$$

definierten neuen Koordinaten

$$u, y_1, y_2, \ldots, y_{n-1}$$

einführen. Dabei sei das System $(u, y_1, y_2, \ldots, y_{n-1})$ eine eineindeutige stetige Transformation des Systems $(x_1, x_2, \ldots, x_n)$ mit überall stetigen partiellen Ableitungen

$$\frac{\partial u}{\partial x_k}, \quad \frac{\partial y_j}{\partial x_k} \qquad (j = 1, 2, \ldots, n-1; \quad k = 1, 2, \ldots, n).$$

Die gemeinsame Dichte der neuen Zufallsvariablen $(U, Y_1, \ldots, Y_{n-1})$ hat die Form

$$f(x_1, Q) f(x_2, Q) \cdots f(x_n, Q) \cdot |J|; \tag{13.4.3}$$

dabei sind die Zufallsvariablen $x_1, x_2, \ldots, x_n$ Funktionen der Zufallsvariablen $u, y_1, y_2, \ldots, y_{n-1}$, und J ist die Funktionaldeterminante der zu (13.4.2) inversen Transformation. Wir bezeichnen mit $g(u, Q)$ die Dichte der Zufallsvariablen U.

Wir betrachten die bedingte Verteilung der Zufallsvariablen $(Y_1, Y_2, \ldots, Y_{n-1})$ unter der Bedingung, daß U einen gegebenen Wert u annimmt. Die Dichte dieser Verteilung wollen wir mit

$$h(y_1, y_2, \ldots, y_{n-1} \mid u, Q) \tag{13.4.4}$$

bezeichnen. Aus (13.4.3) folgt unter Berücksichtigung einer zu (2.7.6) analogen Formel die Gleichung

$$f(x_1, Q) f(x_2, Q) \cdots f(x_n, Q) \cdot |J| = g(u, Q) h(y_1, y_2, \ldots, y_{n-1} \mid u, Q). \tag{13.4.5}$$

Wir wählen die durch (13.4.2) bestimmte Stichprobenfunktion U als Schätzfunktion für den Parameter Q. Wenn in einem gewissen Intervall der Werte von Q die partiellen Ableitungen

$$\frac{\partial f}{\partial Q}, \quad \frac{\partial g}{\partial Q}, \quad \frac{\partial h}{\partial Q}$$

existieren und die Ungleichungen

$$\left|\frac{\partial f}{\partial Q}\right| < F_0(x), \quad \left|\frac{\partial g}{\partial Q}\right| < G_0(u), \quad \left|\frac{\partial h}{\partial Q}\right| < H_0(y_1, y_2, \ldots, y_{n-1}, u)$$

erfüllen, wobei die Integrale

$$\int_{-\infty}^{\infty} F_0(x)\,dx, \quad \int_{-\infty}^{\infty} G_0(u)\,du, \quad \int_{-\infty}^{\infty} u\,G_0(u)\,du$$

und

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_0(y_1, y_2, \ldots, y_{n-1}, u)\,dy_1 \cdots dy_{n-1}$$

endlich sind, nennen wir U eine *reguläre* Schätzfunktion des Parameters Q. Im folgenden wollen wir uns auf reguläre Schätzfunktionen beschränken.

Jetzt können wir auch den Begriff einer erschöpfenden Schätzfunktion definieren.

Definition 13.4.1. Es sei D die Klasse der Verteilungsfunktionen $F(x, Q)$. Eine Stichprobenfunktion U ist eine *erschöpfende Schätzfunktion des Parameters Q in der Klasse D*, wenn für jede Verteilungsfunktion $F(x, Q)$ aus der Klasse D die Dichte (13.4.4) der gemeinsamen bedingten Verteilung der Zufallsvariablen $(Y_1, Y_2, \ldots, Y_{n-1})$ unter der Bedingung $U = u$ für $g(u, Q) > 0$ von Q unabhängig ist.

Im weiteren werden wir kurz von einer *erschöpfenden Schätzfunktion* sprechen, während wir die Klasse D, in der die Schätzfunktion erschöpfend ist, nur dann besonders kennzeichnen werden, wenn sich das nicht von selbst aus dem Text ergibt.

Daher hat die Likelihood-Funktion L, wenn die Schätzfunktion U erschöpfend ist, die Form

$$L = g(u, Q)\,h(y_1, y_2, \ldots, y_{n-1} \,|\, u). \tag{13.4.6}$$

Wenn U eine erschöpfende Schätzfunktion ist, so ist für einen festen Wert u die bedingte Verteilung in Punkten E auf der Hyperfläche mit der Gleichung $U = u$ konzentriert. Da die bedingte Wahrscheinlichkeit auf dieser Hyperfläche von Q unabhängig ist, liefert die Präzisierung der Lage des Punktes E auf dieser Hyperfläche keine neuen Informationen über den Parameter Q. Mit anderen Worten: Ist U eine erschöpfende Stichprobenfunktion, so ist die bedingte Wahrschein-

lichkeit irgendeiner anderen Stichprobenfunktion U_1 unter der Bedingung $U = u$ von Q unabhängig; demzufolge ergeben sich daraus ebenfalls keine neuen Informationen über Q.

Wir bemerken, daß die hier nachgewiesene Eigenschaft einer erschöpfenden Stichprobenfunktion FISHER zur Definition dieses Begriffes gedient hat und daß sie auch die Bezeichnung *erschöpfende Stichprobenfunktion* erklärt. Die hier gegebene Definition 13.4.1 ist, wie NEYMAN [3] und DARMOIS [1] gezeigt haben, der Fisherschen Definition gleichwertig.

Falls eine Stichprobenfunktion erschöpfend ist, hängt, wie aus (13.4.6) hervorgeht, die Maximum-Likelihood-Abschätzung des Parameters Q nur von dem Wert der Variablen U ab, während er von den speziellen Werten der Variablen $X_1, X_2, \ldots, X_n$, deren Funktion U ist, unabhängig ist.

Beispiel 13.4.1. Das Merkmal X einer Gesamtheit sei normal $N(m; \sigma)$ verteilt, wobei der Mittelwert m unbekannt sei. Als Schätzfunktion U des Parameters wählen wir

$$U = \frac{1}{n} \sum_{k=1}^{n} X_k = \bar{X}.$$

Dabei mögen die Variablen X_k voneinander unabhängig sein und dieselbe Verteilung $N(m; \sigma)$ haben. Die Schätzfunktion U ist in jedem Intervall $a < m < b$ regulär. Aus Formel (9.5.10) folgt

$$L = f(x_1, m)\, f(x_2, m) \cdots f(x_n, m) = 2\, \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\, e^{-\frac{n}{2}\frac{(\bar{x}-m)^2}{\sigma^2}}\; \frac{n^{\frac{n-1}{2}} \exp\left(-\dfrac{n s^2}{2\sigma^2}\right) s^{n-2}}{(2\sigma^2)^{\frac{n-1}{2}}\, \Gamma\left(\dfrac{n-1}{2}\right)}. \tag{13.4.7}$$

Wie man sieht, kann man die Likelihood-Funktion L als Produkt darstellen; dabei ist der zweite Faktor auf der rechten Seite die Dichte der Zufallsvariablen S (siehe 9.5). Dieser Faktor ist vom Parameter m unabhängig. Die Schätzfunktion U ist also erschöpfend.

Die hier für stetige Zufallsvariable eingeführten Begriffe lassen sich ohne Schwierigkeiten auf diskrete Zufallsvariable übertragen.

Der Begriff der erschöpfenden Schätzfunktion läßt sich für den Fall verallgemeinern, daß mehrere Parameter unbekannt sind (siehe CRAMÉR [2], SCHMETTERER [1], BARANKIN und KATZ [1] sowie die Aufgaben 13.10.11 bis 13.10.14). Tiefergehende und allgemeinere Grundlagen der Theorie der erschöpfenden Schätzfunktionen findet man in der Arbeit von HALMOS und SAVAGE [1].

13.5. Wirksamste Schätzfunktionen

A. Eine wichtige Eigenschaft einer Schätzfunktion, die anschaulicher als die ist, erschöpfend zu sein, stellt die *Wirksamkeit* dar, die ebenfalls von FISHER [2], [6] eingeführt wurde.

Eine erwartungstreue Schätzfunktion ist eine Zufallsvariable, deren Erwartungswert gleich dem unbekannten Parameter ist. Nun ist es einleuchtend, daß die möglichen Werte einer Schätzfunktion sich desto näher um ihren Erwartungswert häufen, je kleiner die Dispersion dieser Schätzfunktion ist. Um so größer ist dann die Wahrscheinlichkeit dafür, daß die Abschätzung des unbekannten Parameters nahe dem wirklichen Parameterwert liegt. Es ist also angebracht zu fordern, daß die Dispersion der Schätzfunktion möglichst klein sei. Diese Eigenschaft kommt den wirksamsten Schätzfunktionen zu.

Zuerst befassen wir uns mit der Herleitung einer wichtigen Ungleichung, die nach RAO-CRAMÉR (RAO [1], CRAMÉR [2]) benannt ist. Wir wollen sie gesondert für stetige und für diskrete Zufallsvariable formulieren.

Die Ungleichung von RAO-CRAMÉR. *Es sei U eine reguläre erwartungstreue Schätzfunktion des Parameters Q in einfachen Stichproben vom Umfang n.*

Ist das Merkmal X stetig mit der Dichte $f(x,Q)$ verteilt, so gilt

$$D^2(U) = E[(U-Q)^2] \geqq \frac{1}{n\int\limits_{-\infty}^{\infty}\left[\frac{\partial \log f(x,Q)}{\partial Q}\right]^2 f(x,Q)\,dx}. \tag{13.5.1}$$

Ist das Merkmal X diskret mit den Sprungstellen $x_1, x_2, \ldots$ und der Wahrscheinlichkeitsfunktion $P(X = x_i) = p_i(Q)$ verteilt, so gilt

$$D^2(U) = E[(U-Q)^2] \geqq \frac{1}{n\sum\limits_i \left(\frac{d \log p_i(Q)}{dQ}\right)^2 p_i(Q)}. \tag{13.5.1'}$$

Beweis. Wir beschränken uns auf den Beweis der Formel (13.5.1). Aus unseren Voraussetzungen über die Regularität (siehe 13.4) der Schätzfunktion U folgt, daß man die Identität

$$\int\limits_{-\infty}^{\infty} f(x,Q)\,dx = \int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty} h(y_1,\ldots,y_{n-1}\,|\,u,Q)\,dy_1\cdots dy_{n-1} = 1$$

unter dem Integralzeichen nach Q differenzieren kann. Wir erhalten

$$\int\limits_{-\infty}^{\infty}\frac{\partial \log f(x,Q)}{\partial Q} f(x,Q)\,dx \tag{13.5.2}$$

$$= \int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty}\frac{\partial \log h(y_1,\ldots,y_{n-1}\,|\,u,Q)}{\partial Q} h(y_1,\ldots,y_{n-1}\,|\,u,Q)\,dy_1\cdots dy_{n-1} = 0.$$

Wenn wir die logarithmische Ableitung auf beiden Seiten der Gleichung (13.4.5) bilden [dabei betrachten wir ihre linke Seite als Funktion der Zufallsvariablen x_k $(k = 1, 2, \ldots, n)$], so erhalten wir

$$\sum_{k=1}^{n} \frac{\partial \log f(x_k, Q)}{\partial Q} = \frac{\partial \log g(u, Q)}{\partial Q} + \frac{\partial \log h(y_1, \ldots, y_{n-1} | u, Q)}{\partial Q}. \tag{13.5.3}$$

Wir quadrieren beide Seiten der Gleichung (13.5.3) und multiplizieren sie mit der Gleichung (13.4.5). Es ergibt sich

$$\left\{\sum_{k=1}^{n} \left[\frac{\partial \log f(x_k, Q)}{\partial Q}\right]^2 + \sum_{\substack{l,j \\ l \neq j}} \frac{\partial \log f(x_l, Q)}{\partial Q} \frac{\partial \log f(x_j, Q)}{\partial Q}\right\} f(x_1, Q) \cdots f(x_n, Q)$$

$$= \left\{\left[\frac{\partial \log g}{\partial Q}\right]^2 + \left[\frac{\partial \log h}{\partial Q}\right]^2 + 2 \frac{\partial \log g}{\partial Q} \frac{\partial \log h}{\partial Q}\right\} g(u, Q) h(y_1, \ldots, y_{n-1} | u, Q). \tag{13.5.4}$$

Es gilt nun

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{k=1}^{n} \left[\frac{\partial \log f(x_k, Q)}{\partial Q}\right]^2 f(x_1, Q) \cdots f(x_n, Q)\, dx_1 \cdots dx_n$$

$$= n \int_{-\infty}^{\infty} \left[\frac{\partial \log f(x, Q)}{\partial Q}\right]^2 f(x, Q)\, dx. \tag{13.5.5}$$

Die linke Seite der Gleichung (13.5.2) ergibt

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{\substack{l,j \\ l \neq j}} \frac{\partial \log f(x_l, Q)}{\partial Q} \frac{\partial \log f(x_j, Q)}{\partial Q} f(x_1, Q) \cdots f(x_n, Q)\, dx_1 \cdots dx_n = 0, \tag{13.5.6}$$

während aus der rechten Seite von (13.5.2)

$$\int_{-\infty}^{\infty} g(u, Q) \frac{\partial \log g(u, Q)}{\partial Q}\, du$$

$$\times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial \log h}{\partial Q} h(y_1, \ldots, y_{n-1} | u, Q)\, dy_1 \cdots dy_{n-1} = 0 \tag{13.5.7}$$

folgt. Wenn wir beide Seiten der Gleichung (13.5.4) integrieren und die Formeln (13.5.5) bis (13.5.7) beachten, erhalten wir

$$n\int\limits_{-\infty}^{\infty}\left[\frac{\partial \log f(x, Q)}{\partial Q}\right]^2 f(x,Q)\,dx = \int\limits_{-\infty}^{\infty}\left[\frac{\partial \log g(u, Q)}{\partial Q}\right]^2 g(u, Q)\,du$$

$$+\int\limits_{-\infty}^{\infty} g(u, Q)\,du \int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty}\left(\frac{\partial \log h}{\partial Q}\right)^2 h(y_1, \ldots, y_{n-1}\,|\,u, Q)\,dy_1\cdots dy_{n-1}$$

$$\geqq \int\limits_{-\infty}^{\infty}\left[\frac{\partial \log g(u, Q)}{\partial Q}\right]^2 g(u, Q)\,du. \tag{13.5.8}$$

Nach Voraussetzung ist die Stichprobenfunktion U eine erwartungstreue Schätzfunktion für den Parameter Q; also ist

$$E(U) = \int\limits_{-\infty}^{\infty} u g(u, Q)\,du = Q. \tag{13.5.9}$$

Außerdem folgt aus der vorausgesetzten Regularität die Differenzierbarkeit der Gleichung (13.5.9) in bezug auf Q unter dem Integralzeichen. Wir erhalten somit

$$1 = \int\limits_{-\infty}^{\infty} u\,\frac{\partial g(u, Q)}{\partial Q}\,du = \int\limits_{-\infty}^{\infty}(u - Q)\,\frac{\partial g(u, Q)}{\partial Q}\,du\text{[1]}$$

$$= \int\limits_{-\infty}^{\infty}(u - Q)\sqrt{g(u, Q)}\,\frac{\partial \log g(u, Q)}{\partial Q}\sqrt{g(u, Q)}\,du, \tag{13.5.10}$$

also

$$1 = \left[\int\limits_{-\infty}^{\infty}(u - Q)\sqrt{g(u, Q)}\,\frac{\partial \log g(u, Q)}{\partial Q}\sqrt{g(u, Q)}\,du\right]^2.$$

---

[1]) Diese Umformung ergibt sich daraus, daß $\int\limits_{-\infty}^{\infty} g(u, Q)\,du = 1$ ist und daß man dieses Integral unter dem Integralzeichen differenzieren kann, woraus

$$\int\limits_{-\infty}^{\infty}\frac{\partial g(u, Q)}{\partial Q}\,du = 0$$

folgt.

Nach der Schwarzschen Ungleichung (siehe etwa FICHTENHOLZ [1]) erhalten wir die Ungleichung

$$1 \leqq \int_{-\infty}^{\infty} (u - Q)^2 g(u, Q)\, du \int_{-\infty}^{\infty} \left[\frac{\partial \log g(u, Q)}{\partial Q}\right]^2 g(u, Q)\, du. \tag{13.5.11}$$

Wir schreiben (13.5.11) in der Form

$$D^2(U) = \int_{-\infty}^{\infty} (u - Q)^2 g(u, Q)\, du \geqq \frac{1}{\int_{-\infty}^{\infty} \left[\frac{\partial \log g(u, Q)}{\partial Q}\right]^2 g(u, Q)\, du}.$$

Unter Berücksichtigung von (13.5.8) erhalten wir

$$D^2(U) \geqq \frac{1}{n \int_{-\infty}^{\infty} \left[\frac{\partial \log f(x, Q)}{\partial Q}\right]^2 f(x, Q)\, dx}.$$

Die Ungleichung (13.5.1) ist damit bewiesen.

Nach dieser Ungleichung ist die Dispersion der Schätzfunktion U eines Parameters Q nicht kleiner als eine gewisse feste Größe, die durch n und die Verteilung der Gesamtheit bestimmt wird.

B. Wir wollen nun den Begriff der wirksamsten Schätzfunktionen definieren.

Definition 13.5.1. Eine erwartungstreue Schätzfunktion U des Parameters Q heißt *wirksamste Schätzfunktion für die Klasse D von Verteilungsfunktionen*, wenn für jede Verteilungsfunktion $F(x, Q)$ aus der Klasse D die Varianz $D^2(U)$ den durch die Formeln (13.5.1) bzw. (13.5.1′) gegebenen minimalen Wert erreicht.

Im weiteren werden wir häufig kurz *wirksamste Schätzfunktion* sagen.

Aus dieser Definition folgt, daß, allgemein ausgedrückt, die Werte einer erwartungstreuen wirksamsten Schätzfunktion so dicht wie nur möglich um den Wert $E(U) = Q$ gehäuft liegen; die Wahrscheinlichkeit, daß der beobachtete Wert u dieser Schätzfunktion nahe am wirklichen Wert des unbekannten Parameters Q liegt, ist also so groß wie nur möglich.

Satz 13.5.1. *Eine erwartungstreue Schätzfunktion U des Parameters Q ist genau dann eine wirksamste Schätzfunktion, wenn*

a) *die Schätzfunktion erschöpfend ist,*

b) *ihre Dichte* $g(u, Q)$ *für* $g(u, Q) > 0$ *fast überall*[1]) *die Gleichung*

$$\frac{\partial \log g(u, Q)}{\partial Q} = c(u - Q) \tag{13.5.12}$$

erfüllt, wobei die Zahl c *von* u *unabhängig ist (sie kann aber von* Q *abhängen).*

Beweis. Wir wollen zuerst zeigen, daß diese Bedingungen hinreichend sind. Es seien also die Bedingungen a) und b) erfüllt. Nach a) ist die bedingte Dichte h [siehe (13.4.4)] von Q unabhängig. (13.5.8) nimmt also, falls wir die Bedingung b) berücksichtigen, die Form

$$n \int_{-\infty}^{\infty} \left[\frac{\partial \log f(x, Q)}{\partial Q}\right]^2 f(x, Q)\, dx = c^2 \int_{-\infty}^{\infty} (u - Q)^2 g(u, Q)\, du = c^2 D^2(U)$$

an, während (13.5.10) die Gestalt

$$1 = c \int_{-\infty}^{\infty} (u - Q)^2 g(u, Q)\, du = c D^2(U)$$

bekommt. Eliminieren wir aus diesen beiden Gleichungen c, so erhalten wir

$$D^2(U) = \frac{1}{n \int\limits_{-\infty}^{\infty} \left[\frac{\partial \log f(x, Q)}{\partial Q}\right]^2 f(x,Q)\, dx}. \tag{13.5.13}$$

Nun wollen wir die Notwendigkeit der Bedingungen a) und b) nachweisen. Wir nehmen also an, die Gleichung (13.5.13) sei erfüllt. Aus (13.5.8) folgt dann, daß die bedingte Dichte h von Q unabhängig ist, da anderenfalls wegen (13.5.11) die Ungleichung

$$D^2(U) > \frac{1}{n \int\limits_{-\infty}^{\infty} \left[\frac{\partial \log f(x, Q)}{\partial Q}\right]^2 f(x, Q)\, dx}$$

gelten würde, was (13.5.13) widerspricht. Damit ist gezeigt, daß die Schätzfunktion U erschöpfend ist, daß also die Bedingung a) erfüllt ist.

Wir schreiben jetzt (13.5.8) in der Gestalt

$$\frac{1}{D^2(U)} = \int_{-\infty}^{\infty} \left[\frac{\partial \log g(u, Q)}{\partial Q}\right]^2 g(u, Q)\, du.$$

---

[1]) Das heißt mit Ausnahme einer Menge, deren Lebesguesches Maß gleich Null ist.

Dieser Ausdruck ist identisch mit der Formel (13.5.11), in der das Zeichen $\leqq$ durch das Gleichheitszeichen ersetzt wurde. Wie man weiß, ist dafür, daß in der Schwarzschen Ungleichung das Gleichheitszeichen gilt, die Existenz zweier nicht gleichzeitig verschwindender Konstanten w und v notwendig, derart, daß die Gleichung

$$\int_{-\infty}^{\infty} \left[w \cdot (u - Q) \sqrt{g(u, Q)} + v \frac{\partial \log g(u, Q)}{\partial Q} \sqrt{g(u, Q)} \right]^2 du = 0 \quad (13.5.14)$$

gilt. Der Ausdruck unter dem Integralzeichen ist nicht negativ. Die Gleichung (13.5.14) gilt also nur dann, wenn fast überall

$$w \cdot (u - Q) \sqrt{g(u, Q)} + v \frac{\partial \log g(u, Q)}{\partial Q} \sqrt{g(u, Q)} = 0 \quad (13.5.15)$$

ist. Da der Ausdruck $(u - Q)\sqrt{g(u, Q)}$ nicht fast überall verschwinden kann (die Dispersion der Schätzfunktion U wäre dann gleich 0), gilt also für $g(u, Q) > 0$ die Gleichung (13.5.15) nur dann, wenn die Gleichheit (13.5.12) gilt. Die Bedingung b) ist also ebenfalls erfüllt.

C. Es sei U eine erwartungstreue wirksamste Schätzfunktion des Parameters Q, und U_1 sei eine andere erwartungstreue Schätzfunktion dieses Parameters. Als Maß für die *Wirksamkeit* der Schätzfunktion U_1 kann der Ausdruck

$$e = \frac{D^2(U)}{D^2(U_1)} \quad (13.5.16)$$

dienen. Für erwartungstreue wirksamste Schätzfunktionen ist $e = 1$.

Falls die Dispersionen der verglichenen Schätzfunktionen umgekehrt proportional zu n sind, können wir den Wert e folgendermaßen interpretieren. Durch eine wirksamste Schätzfunktion U erhalten wir aus einer Stichprobe vom Umfang n eine Abschätzung des unbekannten Parameters Q von derselben Güte (vom Standpunkt der Wirksamkeit betrachtet) wie mit Hilfe der Schätzfunktion U_1 aus einer $\frac{n}{e}$ Elemente zählenden Stichprobe.

Praktisch ist es angebracht, wirksamste Schätzfunktionen zu verwenden, da sie am ökonomischsten sind; sie existieren aber selten.

Beispiel 13.5.1. Wir zeigten im Beispiel 13.4.1, daß die einfache Stichprobenfunktion $U = \bar{X}$ eine erschöpfende Schätzfunktion für den Parameter m einer Normalverteilung $N(m; \sigma)$ ist, falls die Standardabweichung bekannt ist. Wir zeigen jetzt, daß diese Schätzfunktion auch eine wirksamste ist.

Es genügt zu beweisen, daß die Bedingung (13.5.12) erfüllt ist. Tatsächlich ergibt die Formel (13.4.7)

$$g(\bar{x}, m) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n}{2}\frac{(\bar{x}-m)^2}{\sigma^2}\right), \tag{13.5.17}$$

woraus

$$\frac{\partial \log g(\bar{x}, m)}{\partial m} = \frac{n}{\sigma^2}(\bar{x}-m) \tag{13.5.18}$$

folgt. Somit ist der Stichprobenmittelwert, wenn die Stichprobe einfach ist und aus einer normalen Gesamtheit stammt, eine wirksamste Schätzfunktion des unbekannten Gesamtheitsmittelwerts.

Wie man aus (13.5.17) ersieht, gilt

$$D^2(\bar{X}) = \frac{\sigma^2}{n}. \tag{13.5.19}$$

Dies ist die minimale Dispersion einer regulären und erwartungstreuen Schätzfunktion für den Parameter m einer normalen Gesamtheit in einfachen Stichproben vom Umfang n.

Beispiel 13.5.2. Das Merkmal X einer Gesamtheit sei normal verteilt mit der Dichte

$$f(x, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \tag{13.5.20}$$

Dabei sei m bekannt und σ unbekannt.

Mit Hilfe einer einfachen Stichprobe vom Umfang n wollen wir σ^2 abschätzen, indem wir als Abschätzung den beobachteten Wert s_0^2 der Schätzfunktion

$$S_0^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2 \tag{13.5.21}$$

wählen. Diese Schätzfunktion ist in einem beliebigen Intervall $a < \sigma^2 < b$ $(a > 0)$ regulär. Wie bekannt ist (siehe 13.3), ist S_0^2 eine erwartungstreue Schätzfunktion für die Dispersion σ^2. Wir zeigen, daß sie für die Klasse der Verteilungsfunktionen von Zufallsvariablen mit Normalverteilung auch eine wirksamste ist.

So folgt aus Formel (9.4.9)

$$g(s_0^2, \sigma^2) = \frac{n^{\frac{n}{2}} s_0^{n-2} \exp\left(-\frac{n s_0^2}{2\sigma^2}\right)}{(2\sigma^2)^{\frac{n}{2}}\, \Gamma\left(\frac{n}{2}\right)} \tag{13.5.22}$$

und daraus

$$\frac{\partial \log g(s_0^2, \sigma^2)}{\partial \sigma^2} = \frac{n}{2\sigma^4}(s_0^2 - \sigma^2). \tag{13.5.23}$$

Die Bedingung (13.5.12) ist also erfüllt. Andererseits haben wir

$$L = f(x_1, \sigma^2) f(x_2, \sigma^2) \cdots f(x_n, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left[-\frac{\sum_{k=1}^{n}(x_k - m)^2}{2\sigma^2}\right]$$

$$= \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left(-\frac{n s_0^2}{2\sigma^2}\right) = g(s_0^2, \sigma^2) \frac{\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}}\, n^{\frac{n}{2}}\, s_0^{n-2}}. \tag{13.5.24}$$

Wie wir sehen, ist der zweite Faktor auf der rechten Seite der Formel (13.5.24) von σ^2 unabhängig.

Die Bedingungen a) und b) sind also erfüllt, und S_0^2 ist eine wirksamste Schätzfunktion für den Parameter σ^2.

Die minimale Dispersion für reguläre und erwartungstreue Schätzfunktionen des Parameters σ^2 finden wir aus den Formeln (13.5.13) und (13.5.20):

$$D^2(S_0^2) = \frac{1}{n\int\limits_{-\infty}^{\infty}\left[\frac{\partial \log f(x, \sigma^2)}{\partial \sigma^2}\right]^2 f(x, \sigma^2)\, dx}$$

$$= \frac{1}{n\int\limits_{-\infty}^{\infty}\left[\frac{(x-m)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right]^2 f(x, \sigma^2)\, dx} = \frac{2\sigma^4}{n}. \tag{13.5.25}$$

Beispiel 13.5.3. Das Merkmal X einer Gesamtheit sei normal $N(m;\sigma)$ verteilt. Wir wählen als Abschätzung der Dispersion σ^2 den beobachteten Wert u der Schätzfunktion

$$U = \frac{n}{n-1} S^2 = \frac{\sum_{k=1}^{n}(X_k - \bar{X})^2}{n-1} \tag{13.5.26}$$

in einer einfachen Stichprobe vom Umfang n.

Diese Schätzfunktion U ist erwartungstreu (siehe 13.3) und in einem beliebigen Intervall $a < \sigma^2 < b$ $(a > 0)$ regulär. Die Schätzfunktion ist jedoch nicht erschöpfend, folglich erst recht keine wirksamste. Dies folgt auch unmittelbar aus (13.4.7), da die Standardabweichung σ doch in beiden Faktoren auf der rechten Seite dieser Formel vorkommt. Leicht sieht man, daß

$$D^2\left(\frac{n}{n-1} S^2\right) = \frac{2\sigma^4}{n-1}$$

ist. Die Wirksamkeit dieser Schätzfunktion beträgt nach Formel (13.5.16)

$$e = \frac{D^2(S_0^2)}{D^2\left(\frac{n}{n-1} S^2\right)} = \frac{n-1}{n} < 1.$$

Beispiel 13.5.4. Wir setzen dieselben Bedingungen wie im Beispiel 13.5.3 voraus und wollen die Standardabweichung σ mit Hilfe der Schätzfunktion

$$U = \frac{\sqrt{n}\,\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)} S_0 \tag{13.5.27}$$

abschätzen.

Wie schon bekannt [siehe Beispiel 13.3.3, Formel (13.3.8)], ist U eine erwartungstreue Schätzfunktion des Parameters σ. Mit einer ähnlichen Überlegung wie bei der Stichprobenfunktion S_0^2 können wir erkennen, daß U eine erschöpfende Schätzfunktion für den Parameter σ ist. Jedoch ist U für diesen Parameter keine wirksamste Schätzfunktion. Wir haben nämlich

$$g(u, \sigma) = \frac{2K_n^{n-1}}{\sigma^n} u^{n-1} \exp\left[-\frac{K_n^2 u^2}{\sigma^2}\right] \tag{13.5.28}$$

mit

$$K_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

Daraus erhalten wir

$$\frac{\partial \log g(u, \sigma)}{\partial \sigma} = \frac{2K_n}{\sigma^3}\left(u^2 - \frac{n\sigma^2}{2K_n}\right).$$

Die Bedingung (13.5.12) ist also nicht erfüllt.

Den minimalen Dispersionswert einer regulären und erwartungstreuen Schätzfunktion für den Parameter σ finden wir aus Formel (13.5.1). Er beträgt

$$\frac{1}{n\int\limits_{-\infty}^{\infty}\left[\frac{(x-m)^2}{\sigma^3} - \frac{1}{\sigma}\right]^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx} = \frac{\sigma^2}{2n}. \tag{13.5.29}$$

Man kann zeigen, daß die Dispersion der Schätzfunktion U, die durch die Formel (13.5.27) definiert wurde, wie auch zu erwarten war, größer als $\dfrac{\sigma^2}{2n}$ ist.

Beispiel 13.5.5. Das Merkmal X einer Gesamtheit sei normal $N(m; \sigma)$ verteilt. Dabei sei m bekannt, während σ ein unbekannter Parameter sei. Als Abschätzung dieses Parameters wählen wir in einer Stichprobe vom Umfang n die mit $\sqrt{0{,}5\pi}$ multiplizierte mittlere Abweichung, d. h. den beobachteten Wert u der durch

$$U = \frac{1}{n}\sqrt{0{,}5\pi}\sum_{k=1}^{n}|X_k - m| \tag{13.5.30}$$

definierten Schätzfunktion U. Die Variablen X_k seien dabei unabhängig und mögen dieselbe Verteilung haben wie die Variable X. Wir untersuchen die Wirksamkeit dieser Schätzfunktion U. Zuerst wollen wir nachprüfen, ob U für den Parameter σ erwartungstreu ist.

Wir haben

$$E(|X-m|) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x-m|\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx$$

$$= \frac{2}{\sigma\sqrt{2\pi}} \int_{m}^{\infty} (x-m)\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx = \sigma\sqrt{\frac{2}{\pi}},$$

also

$$E(U) = \frac{1}{n}\sqrt{\frac{\pi}{2}} \sum_{k=1}^{n} E(|X_k-m|) = \sqrt{\frac{\pi}{2}}\, E(|X-m|) = \sigma. \tag{13.5.31}$$

Ähnlich erhalten wir

$$D^2(U) = \frac{\pi-2}{2n}\,\sigma^2. \tag{13.5.32}$$

Diese Schätzfunktion U ist aber keine wirksamste. Ihre Wirksamkeit e beträgt nämlich

$$e = \frac{\sigma^2}{2n} : \frac{\pi-2}{2\pi}\,\sigma^2 = \frac{1}{\pi-2} \approx 0{,}876. \tag{13.5.33}$$

D. Wir wollen jetzt wirksamste Schätzfunktionen in Stichproben aus Gesamtheiten betrachten, in welchen das untersuchte Merkmal eine diskrete Zufallsvariable ist.

Beispiel 13.5.6. Das Merkmal X einer Gesamtheit sei binomial verteilt, d. h., es gelte

$$p_k = P(X=k) = \binom{r}{k} p^k q^{r-k}, \quad q = 1-p, \quad 0 < p < 1 \quad (k = 0, 1, \ldots, r). \tag{13.5.34}$$

Dabei sei p der unbekannte Parameter, den wir auf Grund einer aus dieser Gesamtheit erhobenen einfachen Stichprobe vom Umfang n abschätzen wollen.

Wir berechnen nach (13.5.1′) die minimale Dispersion für eine reguläre und erwartungstreue Schätzfunktion U des Parameters p:

$$\frac{1}{D^2(U)} = n\sum_{k=0}^{r} \left(\frac{d\log p_k}{dp}\right)^2 p_k = n\sum_{k=0}^{r} \left(\frac{k}{p} - \frac{r-k}{1-p}\right)^2 p_k$$

$$= n\sum_{k=0}^{r} \left(\frac{k-pr}{pq}\right)^2 p_k = \frac{n}{p^2q^2}\, D^2(X) = \frac{n}{p^2q^2}\, pqr = \frac{nr}{pq}. \tag{13.5.35}$$

Als Abschätzung für den Parameter p wählen wir den beobachteten Wert der Schätzfunktion

$$U = \frac{\overline{X}}{r} = \frac{1}{rn} \sum_{j=1}^{n} X_j, \tag{13.5.36}$$

wobei die Zufallsvariablen X_j unabhängig seien und dieselbe durch Formel (13.5.34) bestimmte Verteilung haben.

Wir haben

$$E(U) = \frac{nrp}{nr} = p \qquad D^2(U) = \frac{npqr}{r^2 n^2} = \frac{pq}{nr}. \tag{13.5.37}$$

Daraus folgt, daß die Stichprobenfunktion U eine erwartungstreue und wirksamste Schätzfunktion ist.

Wir betrachten den Spezialfall $r = 1$, d. h., die Variable X habe eine Null-Eins-Verteilung mit $P(X = 1) = p$, $P(X = 0) = q$. Als Abschätzung des unbekannten Parameters wählen wir den Wert $u = p^* = \bar{x}$ der Stichprobenfunktion aus der Formel (13.5.36). Dies ist der beobachtete Wert der wirksamsten Schätzfunktion für den Parameter p.

In einer Gesamtheit mögen sich fehlerlose und fehlerhafte Exemplare befinden, der Anteil p der fehlerhaften Stücke sei unbekannt. Dem Auftreten eines fehlerhaften Stückes ordnen wir die Zahl 1, dem eines fehlerlosen die Zahl 0 zu. Wenn sich in einer Stichprobe vom Umfang n nun n_1 fehlerhafte Stücke und $n - n_1$ fehlerlose Stücke befinden, so wählen wir als Abschätzung des Parameters p die Zahl

$$\frac{n_1 \cdot 1 + (n - n_1) \cdot 0}{n} = \frac{n_1}{n};$$

das ist der Anteil der fehlerhaften Stücke in der Stichprobe. Somit kommen wir zu dem Schluß, daß die Häufigkeit einer bestimmten Kategorie von Elementen in einfachen Stichproben eine wirksamste Schätzfunktion für die Häufigkeit dieser Kategorie in der Gesamtheit ist.

Beispiel 13.5.7. Das Merkmal X einer Gesamtheit habe die Poissonsche Verteilung

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (k = 0, 1, 2, \ldots), \tag{13.5.38}$$

wobei der Parameter λ unbekannt sei.

Die minimale Dispersion für eine reguläre und erwartungstreue Schätzfunktion U des Parameters λ in einfachen Stichproben vom Umfang n, die aus dieser Gesamtheit stammen, berechnen wir nach der Formel (13.5.1'):

$$\frac{1}{D^2(U)} = n \sum_{k=0}^{\infty} \left(\frac{d \log p_k}{d\lambda}\right)^2 p_k = n \sum_{k=0}^{\infty} \left(\frac{k}{\lambda} - 1\right)^2 p_k = \frac{n}{\lambda^2} \sum_{k=0}^{\infty} (k - \lambda)^2 p_k$$

$$= \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}. \tag{13.5.39}$$

Als Abschätzung für den Parameter λ wählen wir den beobachteten Wert

$$u = \lambda^* = \bar{x}.$$

Man stellt leicht fest, daß dies der beobachtete Wert einer erwartungstreuen und wirksamsten Schätzfunktion ist. Es gilt also

$$E(U) = \lambda, \qquad D^2(U) = \frac{\lambda}{n}. \tag{13.5.40}$$

E. Wir wollen zum Abschluß über neue Ergebnisse bezüglich der Wirksamkeit der erwartungstreuen Schätzfunktionen sowie über den Satz von BLACKWELL informieren.

Es sei D die betrachtete Klasse der Verteilungsfunktionen $F(x, Q)$. Wir betrachten sämtliche erwartungstreuen Schätzfunktionen des Parameters Q in der Klasse D, die für jede Verteilungsfunktion $F(x, Q)$ aus D eine endliche Varianz haben. Dann ist, wie LEHMANN und SCHEFFÉ [1] gezeigt haben, höchstens eine dieser Schätzfunktionen in der betrachteten Klasse D am wirksamsten. Gewisse Verallgemeinerungen dieses Ergebnisses hat SCHMETTERER [2] angegeben. Das von LEHMANN und SCHEFFÉ behandelte Problem wurde in der Arbeit von STEIN [3] unter einem anderen Gesichtspunkt untersucht.

BLACKWELL [2] hat gezeigt, daß bei Vorliegen einer erwartungstreuen Schätzfunktion V des Parameters Q sowie einer weiteren erschöpfenden Schätzfunktion U desselben Parameters sich eine weitere erwartungstreue Schätzfunktion des Parameters Q konstruieren läßt, die wenigstens ebenso wirksam ist wie V. Dies geht aus dem folgenden Satz von BLACKWELL hervor:

Satz 13.5.2. *Es sei V eine erwartungstreue Schätzfunktion des Parameters Q und U eine erschöpfende Schätzfunktion des Parameters Q. Dann ist die Zufallsvariable $E(V \mid u)$* (siehe 3.6.C) *ebenfalls eine erwartungstreue Schatzfunktion des Parameters Q. Existiert darüber hinaus die Varianz $D^2(V)$, so gilt die Ungleichung*

$$D^2[E(V \mid u)] \leqq D^2(V). \tag{13.5.41}$$

Das Gleichheitszeichen in (13.5.41) *steht nur dann, wenn die Beziehung $E(V \mid u) = V$ mit der Wahrscheinlichkeit* 1 *erfüllt ist.*

Beweis. Aus der Voraussetzung, daß die Schätzfunktion U erschöpfend ist, folgt, daß $E(V \mid u)$ nicht von Q abhängt. Da die Schätzfunktion V erwartungstreu ist, erhalten wir unter Anwendung der Formel (3.6.24')

$$E(V) = E[E(V \mid u)] = Q;$$

daher ist $E(V \mid u)$ eine erwartungstreue Schätzfunktion des Parameters Q.

Wir geben nun den Beweis der Formel (13.5.41) für den Fall, daß alle im Satz auftretenden Zufallsvariablen vom stetigen Typ sind (vgl. 2.7.B).

Es seien $f(v, u)$ bzw. $g(u)$ die Dichtefunktionen der Zufallsvariablen (V, U) und U, und es möge $h(v \mid u)$ die bedingte Dichte der Variablen V unter der Annahme

$U = u$ sein. Wegen (3.6.23′) haben wir

$$D^2(V) = E(V^2) - Q^2 \tag{13.5.42}$$

$$= \int_{-\infty}^{\infty} g(u) \int_{-\infty}^{\infty} v^2 h(v|u)\, dv\, du - \int_{-\infty}^{\infty} g(u) \left[\int_{-\infty}^{\infty} v h(v|u)\, dv\right]^2 du$$

$$+ \int_{-\infty}^{\infty} g(u) \left[\int_{-\infty}^{\infty} v h(v|u)\, dv\right]^2 du - \left[\int_{-\infty}^{\infty} g(u) \int_{-\infty}^{\infty} v h(v|u)\, dv\, du\right]^2$$

$$= A + B,$$

wobei A die Summe aus dem ersten und zweiten Integral auf der rechten Seite von (13.5.42) und B die Summe der beiden anderen Integrale ist.

Wir können also schreiben

$$A = \int_{-\infty}^{\infty} g(u) \left\{ \int_{-\infty}^{\infty} v^2 h(v|u)\, dv - \left[\int_{-\infty}^{\infty} v h(v|u)\, dv\right]^2 \right\} du, \tag{13.5.43}$$

zumal sich B in der Form

$$B = E\{[E(V|u)]^2\} - \{E[E(V|u)]\}^2 = D^2[E(V|u)] \tag{13.5.44}$$

schreiben läßt. Der Ausdruck in der eckigen Klammer auf der rechten Seite von (13.5.43) ist die bedingte Varianz der Zufallsvariablen V unter der Annahme $U = u$; A ist somit nichtnegativ. Aus den Beziehungen (13.5.43) bis (13.5.44) folgt aber

$$D^2(V) = A + B \geqq B = D^2[E(V|u)];$$

die Beziehung (13.5.41) ist somit erfüllt.

Das Gleichheitszeichen in (13.5.41) gilt, wenn $A = 0$, d. h. (vgl. 5.1), wenn die bedingte Verteilung der Zufallsvariablen V unter der Voraussetzung $U = u$ mit der Wahrscheinlichkeit 1 singulär ist. Mit anderen Worten, die Zufallsvariable V nimmt für jeden gegebenen Wert u einen (zufälligen) Wert mit der Wahrscheinlichkeit 1 an. Es gilt daher $E(V|u) = V$ mit der Wahrscheinlichkeit 1.

Barankin [1] erweiterte den Satz von Blackwell auf beliebige zentrale Momente der Schätzfunktionen eines unbekannten Parameters.

F. Der Begriff der wirksamsten Schätzfunktion läßt sich (Cramér [2]) für den Fall der Abschätzung mehrerer unbekannter Parameter erweitern. Es muß dann der Begriff der *gemeinsamen wirksamsten Schätzfunktion* eingeführt werden, wozu eine Verallgemeinerung der Ungleichung von Rao-Cramér benötigt wird. Diese

Wert des Integrals nicht kleiner. Wir haben daher

$$1 - M \leqq n! \int_0^{u_0} \cdots \int_{z_j}^{u_0} \int_{z_{j+1}}^{1} \cdots \int_{z_{n-1}}^{1} dz_n \cdots dz_j \, dz_{j-1} \cdots dz_1. \tag{16.6.13}$$

Nach Berechnung des Integrals auf der rechten Seite von (16.6.13) erhalten wir

$$M \geqq \sum_{k=1}^{j+1} \binom{n}{k-1} u_0^{k-1} (1-u_0)^{n-k+1} = \sum_{i=0}^{j} \binom{n}{i} u_0^i (1-u_0)^{n-i}$$
$$= 1 - I_{u_0}(j+1, n-j).$$

Damit ist die Ungleichung (16.6.8) bewiesen.

Birnbaum [3] hat auch die obere Grenze der Güte dieses Tests angegeben. Mit dem Problem der Güte des Tests D_n^- im Vergleich zur Güte einiger anderer nichtparametrischer Tests beschäftigte sich Chapman [3]. In der gleichen Arbeit empfiehlt Chapman einige neue Gütekriterien für nichtparametrische Tests.

D. Aus der Beziehung (16.6.8) erhält man leicht einen asymptotischen Ausdruck für die untere Grenze des betrachteten Tests bei großen n, wobei sich herausstellt, daß sich diese untere Grenze dann mit Hilfe der Normalverteilung darstellen läßt. Wir wollen jedoch zu dem gleichen Ergebnis nach einem anderen, von Massey [1] angegebenen Verfahren gelangen. Wir bestimmen hierbei die untere Grenze des auf der Stichprobenfunktion $D_n = \sup_{-\infty<x<\infty} |F_0(x) - S_n(x)|$ beruhenden Tests, das gleiche Verfahren läßt sich auf den auf der Stichprobenfunktion D_n^- beruhenden Test anwenden.

Wiederum sei die stetige Verteilungsfunktion $F(x)$ unbekannt. Die Nullhypothese sei $H_0[F(x) \equiv F_0(x)]$, die Alternativhypothese sei $H_3[F(x) \equiv G(x)]$, wobei $F_0(x)$ gegeben ist, $G(x)$ der Beziehung

$$\sup_{-\infty<x<\infty} |F_0(x) - G(x)| = \delta > 0 \tag{16.6.14}$$

genügt und der Wert von $G(x)$ an der Stelle x_0 gegeben ist, für die die Gleichung

$$|F_0(x_0) - G(x_0)| = \delta \tag{16.6.15}$$

gilt. Wir setzen den Kolmogoroffschen λ-Test an und lehnen die Hypothese H_0 zugunsten ihrer Alternativhypothese ab, wenn der Wert $d_n \geqq \lambda_\alpha \sqrt{n}$ beobachtet wird; dabei genügt λ_α der Beziehung

$$P\left(\sqrt{n}\, D_n \geqq \lambda_\alpha\right) = \alpha. \tag{16.6.16}$$

Den Wert von λ_α bestimmen wir aus der Tafel VIII. Die Testgüte M ist durch

$$P\left(\sqrt{n}\, D_n \geqq \lambda_\alpha \,|\, G(x)\right)$$

gegeben. Wir haben

$$M \geqq P\left(\sqrt{n}\,|S_n(x_0) - F_0(x_0)| \geqq \lambda_\alpha \mid G(x)\right)$$

$$= 1 - P\left(F_0(x_0) - \frac{\lambda_\alpha}{\sqrt{n}} < S_n(x_0) < F_0(x_0) + \frac{\lambda_\alpha}{\sqrt{n}} \;\Big|\; G(x)\right).$$

Sodann setzen wir

$$F_0(x_0) = G(x_0) \pm \delta$$

und erhalten nach einfachen Umformungen

$$M \geqq 1 - P\left(\frac{-\lambda_\alpha \pm \delta\sqrt{n}}{\sqrt{G(x_0)[1 - G(x_0)]}} < \frac{[S_0(x_0) - G(x_0)]\sqrt{n}}{\sqrt{G(x_0)[1 - G(x_0)]}} < \frac{\lambda_\alpha \pm \delta\sqrt{n}}{\sqrt{G(x_0)[1 - G(x_0)]}} \;\Big|\; G(x)\right). \tag{16.6.17}$$

Wir bemerken, daß $S_n(x_0)$ bei Gültigkeit der Hypothese H_3 auf Grund des Satzes von Moivre-Laplace die Normalverteilung

$$N\left(G(x_0);\ \sqrt{G(x_0)[1 - G(x_0)]}/\sqrt{n}\right)$$

hat. Wir erhalten daher schließlich aus (16.6.17)

$$M \geqq 1 - \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{t^2}{2}\right) dt$$

mit

$$a = \frac{-\lambda_\alpha \pm \delta\sqrt{n}}{\sqrt{G(x_0)[1 - G(x_0)]}}, \quad b = \frac{\lambda_\alpha \pm \delta\sqrt{n}}{\sqrt{G(x_0)[1 - G(x_0)]}}.$$

E. Wir zeigen, daß der Kolmogoroffsche λ-Test bezüglich der Menge der stetigen Verteilungsfunktionen konsistent ist, die der Beziehung (16.6.14) genügen. In der Tat, die Funktion $G(x)$ möge dieser Beziehung genügen und eine echte Verteilungsfunktion sein; dann haben wir

$$\delta = \sup_{-\infty<x<\infty} |G(x) - F_0(x)| = \sup_{-\infty<x<\infty} |G(x) - S_n(x) + S_n(x) - F_0(x)|$$

$$\leqq \sup_{-\infty<x<\infty} |G(x) - S_n(x)| + \sup_{-\infty<x<\infty} |S_n(x) - F_0(x)|$$

oder

$$\sup_{-\infty<x<\infty} |S_n(x) - F_0(x)| \geqq \delta - \sup_{-\infty<x<\infty} |G(x) - S_n(x)|.$$

Nach dem Satz von Gliwenko (vgl. 10.10) erhalten wir unter Beachtung der Tatsache, daß $G(x)$ eine echte Verteilungsfunktion ist,

$$P\left(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |S_n(x) - F_0(x)| \geqq \delta\right) = 1. \tag{16.6.18}$$

Wir bezeichnen nun mit $\lambda_{n,\alpha}$ eine Zahl, die der Gleichung

$$P\left(\sqrt{n}\, D_n \geqq \lambda_{n,\alpha}\right) = \alpha$$

genügt. Es ist also

$$\lim_{n\to\infty} M = \lim_{n\to\infty} P\left(\sup_{-\infty<x<\infty} |S_n(x) - F_0(x)| \geqq \frac{\lambda_{n,\alpha}}{\sqrt{n}} \,\middle|\, G(x)\right).$$

Aus dem Satz von Kolmogoroff (siehe 10.11) folgt

$$\lim_{n\to\infty} \lambda_{n,\alpha} = \lambda_\alpha.$$

Für hinreichend große n ist daher $\lambda_{n,\alpha}/\sqrt{n} < \delta$. Wegen (16.6.18) ergibt sich

$$\lim_{n\to\infty} M = 1,$$

und der Kolmogoroffsche λ-Test ist somit bezüglich der betrachteten Menge der alternativen Verteilungsfunktionen konsistent.

Nach einem Verfahren, das dem auf den Kolmogoroffschen λ-Test angewandten ähnlich ist, läßt sich für große n ein asymptotischer Ausdruck für die Güte des Tests gewinnen, welcher auf dem Satz von Smirnow (siehe 10.11) beruht, wenn wir die Hypothese $H_0[F_1(x) \equiv F_2(x)]$ im Vergleich zur Alternativhypothese $H_4\left[\sup_{-\infty<x<\infty} |F_1(x) - F_2(x)| = \delta > 0\right]$ testen; dabei sind $F_1(x)$ und $F_2(x)$ unbekannte stetige Verteilungsfunktionen. Durch ähnliche Überlegungen wie vorhin läßt sich zeigen, daß der Smirnowsche λ-Test bezüglich der Menge der stetigen Verteilungsfunktionen konsistent ist, die der folgenden Beziehungen genügen:

$$\sup_{-\infty<x<\infty} |F_1(x) - F_2(x)| = \delta > 0.$$

F. Wir wollen uns nun der Güte des Tests von Wilcoxon-Mann-Whitney zuwenden. Dabei wollen wir die gleichen Bezeichnungen verwenden wie in 12.6.B. Somit sind die Zufallsvariablen X_1 und X_2 unabhängig, wobei X_i $(i = 1, 2)$ eine

stetige Verteilungsfunktion $F_i(x)$ hat. Wir stellen die Nullhypothese $H_0[F_1(x) \equiv F_2(x)]$ auf.

Nunmehr müssen wir die Alternativhypothese zu H_0 formulieren. Die Überlegungen in 12.6.B und insbesondere die Struktur des betrachteten Tests selbst suggerieren die Annahme von H_1 als Alternativhypothese, derzufolge für jedes reelle x, für das $0 < F_1(x) < 1$ gilt, die Ungleichung

$$F_2(x) < F_1(x) \tag{16.6.19}$$

erfüllt ist.

Wir wollen jedoch die allgemeinere Hypothese H_2 betrachten, derzufolge die Ungleichung

$$0 < p = P(X_2 - X_1 < 0) < \frac{1}{2} \tag{16.6.20}$$

gilt.

Daß die Hypothese H_2 allgemeiner ist als die Hypothese H_1, geht aus der Tatsache hervor, daß aus der Gültigkeit der Ungleichung (16.6.19) für jedes x auch die Gültigkeit von (16.6.20) für jedes x folgt. In der Tat, es ist dann (siehe 12.6.B)

$$p = \int_{-\infty}^{\infty} F_2(x)\, dF_1(x) < \int_{-\infty}^{\infty} F_1(x)\, dF_1(x) = \frac{1}{2}.$$

Es liegen n_1 und n_2 unabhängige Beobachtungen der Zufallsvariablen X_1 und X_2 vor, und zwar x_{1l} $(l = 1, 2, \ldots, n_1)$ und x_{2j} $(j = 1, 2, \ldots, n_2)$. Aus diesen Beobachtungswerten berechnen wir den Beobachtungswert u der Stichprobenfunktion U von MANN-WHITNEY und lehnen die Hypothese H_0 zugunsten der Alternativhypothese H_2 ab, wenn $u \leqq u_\alpha$ gilt; hierbei ist α die Sicherheitswahrscheinlichkeit, und u_α genügt der Beziehung

$$\psi = P(U \leqq u_\alpha \mid H_0) \leqq \alpha. \tag{16.6.21}$$

Um die Güte dieses Tests zu ermitteln, beschränken wir uns auf große Stichproben und wenden die von LEHMANN [1] (vgl. 12.6.B) bewiesene Tatsache an, daß die Stichprobenfunktion U asymptotisch normal verteilt ist, und zwar hat die Stichprobenfunktion

$$V = \frac{U - E(U)}{\sqrt{D^2(U)}} \tag{16.6.22}$$

für $n_2 = s n_1$ $(s > 0)$ und $n_1 \to \infty$ die Normalverteilung $N(0; 1)$, wobei $E(U)$ und $D^2(U)$ durch die Formeln (12.6.17) und (12.6.18) gegeben sind. Wir lehnen

somit die Hypothese H_0 ab, wenn der Wert $(u - m_0)/\sigma_0 \leqq v_\alpha$ beobachtet wird. Herbei ist $\Phi(v_\alpha) = \alpha$ und $m_0 = E(U|H_0) = \frac{1}{2}\, n_1 n_2$ sowie. $\sigma_0^2 = D^2(U|H_0)$ $= \frac{1}{12}\, n_1 n_2 (n_1 + n_2 + 1)$. Die Größen u und v_α stehen in folgendem Zusammenhang. Wir setzen $u_\alpha = m_0 - c_\alpha \sigma_0$. Dann haben wir

$$\lim_{n_1 \to \infty} \psi = \alpha, \quad \lim_{n_1 \to \infty} c_\alpha = -v_\alpha .$$

Die Güte dieses Tests erhalten wir somit nach der Formel

$$M = P\left(\frac{U - m_0}{\sigma_0} \leqq v_\alpha \;\middle|\; H_2\right) = \Phi\left(\frac{v_\alpha \sigma_0 + m_0 - E(U)}{\sqrt{D^2(U)}}\right). \tag{16.6.23}$$

Beispiel 16.6.1. Es seien $F_1(x)$ und $F_2(x)$ stetige Verteilungsfunktionen. Mit Hilfe des eben beschriebenen Tests wollen wir die Hypothese $H_0[F_1(x) \equiv F_2(x)]$ im Vergleich zur Alternativhypothese $H_1[F_2(x) \equiv F_1^k(x)]$ überprüfen, wobei k eine gewisse natürliche Zahl größer als 1 ist. Aus den Beziehungen (12.6.8), (12.6.17) und (12.6.18) finden wir

$$p = \frac{1}{k+1}, \qquad E(U) = \frac{n_1 n_2}{k+1},$$

$$D^2(U) = \frac{n_1 n_2}{(k+1)^2}\left(\frac{n_1 - 1}{k+2}\, k + \frac{n_2 - 1}{2k+1}\, k^2 + 1\right).$$

Um zu numerischen Ergebnissen zu gelangen, setzen wir hier $k = 2$, $n_1 = n_2 = 25$ und $\alpha = 0{,}05$. Dann haben wir $m_0 = 312{,}5$ und $\sigma_0^2 = 2656{,}25$. Wir werden somit die Hypothese H_0 zugunsten der Alternativhypothese H_1 ablehnen, wenn der beobachtete Wert u der Ungleichung $(u - 312{,}5)/\sqrt{2656{,}25} \leqq -1{,}64$ genügt oder aber wenn $u \leqq 230$ gilt. Gilt die Alternativhypothese, so ist $E(U) = 208\,\frac{1}{3}$ und $D^2(U) = 2180\,\frac{5}{9}$. Aus (16.6.23) erhalten wir $M = \Phi(0{,}485) = 0{,}6866$.

Hätten wir in diesem Beispiel $n_1 = n_2 = 60$ gewählt, so hätte sich M zu etwa 0,95 ergeben (siehe LEHMANN [3]).

Dem Problem der Güte des Tests von WILCOXON-MANN-WHITNEY und dem Vergleich dieser Güte mit der Güte anderer nichtparametrischer Tests sind Arbeiten von VAN DER VAART [1], VAN DER WAERDEN [2], VAN DANTZIG [1] und LEHMANN [3] gewidmet.

Die Ergebnisse dieser Arbeiten scheinen auszusagen, daß die Güte des Tests von WILCOXON-MANN-WHITNEY besser ist als die Güte der anderen von uns betrachteten nichtparametrischen Tests.

G. Wir wollen nun zeigen, daß bei Verifizierung der Hypothese $H_0[F_1(x) \equiv F_2(xn)]$ im Vergleich zu der durch (16.6.20) gegebenen Alternativhypothese der Test von WILCOXON-MANN-WHITNEY konsistent ist. Beim Beweis werden wir die beiden

nachstehenden Ungleichungen verwenden, die sich unmittelbar aus (12.6.13) und (12.6.14) herleiten lassen:

$$0 \leqq \gamma^2 = \int_{-\infty}^{\infty} F_2^2(x)\, dF_1(x) - p^2 \leqq \int_{-\infty}^{\infty} F_2(x)\, dF_1(x) - p^2 = p - p^2$$
$$= p(1-p),$$
$$0 \leqq \varphi^2 = \int_{-\infty}^{\infty} \big(1 - F_1(x)\big)^2\, dF_2(x) - p^2$$
$$\leqq \int_{-\infty}^{\infty} \big(1 - F_1(x)\big)\, dF_2(x) - p^2 = 1 - (1-p) - p^2 = p(1-p).$$

Aus diesen Ungleichungen und aus (12.6.18) folgt

$$\frac{D^2(U)}{\sigma_0^2} = 12\,\frac{(n_1-1)\varphi^2 + (n_2-1)\gamma^2 + p(1-p)}{n_1+n_2+1}$$
$$\leqq \frac{12\,p(1-p)(n_1+n_2-1)}{n_1+n_2+1} \leqq 12\,p(1-p). \tag{16.6.24}$$

Nun bestimmen wir die Testgüte und erhalten

$$1 - M = P(U > u_\alpha \mid H_2) = P(U > m_0 - c_\alpha \sigma_0 \mid H_2) \tag{16.6.25}$$
$$= P\left(U - n_1 n_2 p > \left(\frac{1}{2} - p\right) n_1 n_2 - c_\alpha \sigma_0 \mid H_2\right).$$

Wir bemerken, daß wegen (16.6.20) die Differenz $\frac{1}{2} - p$ positiv ist. Da ferner c_α beschränkt ist und $n_2 = s n_1$ $(s > 0)$ gilt, ergibt sich für hinreichend große n_1

$$\left(\frac{1}{2} - p\right) n_1 n_2 - c_\alpha \sigma_0 = \left(\frac{1}{2} - p\right) s n_1^2 - \frac{1}{12} c_\alpha n_1 \sqrt{s(2n_1+1)} > 0.$$

Wenden wir nun auf die rechte Seite von (16.6.25) die Tschebyscheffsche Ungleichung an (siehe 3.3), so erhalten wir

$$1 - M \leqq \frac{D^2(U)}{\left(\left(\frac{1}{2} - p\right) n_1 n_2 - c_\alpha \sigma_0\right)^2} = \frac{D^2(U)}{\sigma_0^2}\left[\left(\frac{1}{2} - p\right)\sqrt{\frac{12\, n_1 n_2}{n_1+n_2+1}} - c_\alpha\right]^{-2}.$$

Im Einklang mit (16.6.24) ist der Ausdruck $D^2(U)/\sigma_0^2$ beschränkt. Beachten wir wiederum die Beziehung (16.6.20), die Beziehung $n_2 = s n_1$ sowie die Tatsache,

daß c_α beschränkt ist, so sehen wir, daß der Ausdruck innerhalb der äußeren Klammer auf der rechten Seite der letzten Ungleichung für $n_1 \to \infty$ über alle Grenzen wächst, und es ist somit

$$\lim_{n_1\to\infty} (1 - M) = 0.$$

Schließlich erhalten wir daher für $p < \frac{1}{2}$

$$\lim_{n_1\to\infty} P(U \leqq u_\alpha | H_2) = 1.$$

Damit ist die Konsistenz des Tests von WILCOXON-MANN-WHITNEY bewiesen.

Es sei jedoch erwähnt, daß bei Verifizierung der Hypothese $H_0[F_1(x) \equiv F_2(x)]$ im Vergleich zur Alternativhypothese $H'\left[F_1(x) \not\equiv F_2(x), p \geqq \frac{1}{2}\right]$ der Test von WILCOXON-MANN-WHITNEY nicht konsistent ist. In der Tat, aus (16.6.25) erhalten wir

$$M = P(U \leqq u_\alpha | H') = P\left(U - n_1 n_2 p < -\left[\left(p - \frac{1}{2}\right) n_1 n_2 + c_\alpha \sigma_0\right] H'\right).$$

Da $p \geqq \frac{1}{2}$ und c_α für jedes $\alpha < \frac{1}{2}$ (außer für sehr kleine n_1 und n_2) positiv ist, gilt $\left(p - \frac{1}{2}\right) n_1 n_2 + c_\alpha \sigma_0 > 0$. Aus der Tschebyscheffschen Ungleichung erhalten wir daher

$$M \leqq \frac{D^2(U)}{\left[\left(p - \frac{1}{2}\right) n_1 n_2 + c_\alpha \sigma_0\right]^2} = \frac{D^2(U)}{\sigma_0^2}\left[\left(p - \frac{1}{2}\right)\sqrt{\frac{12 n_1 n_2}{n_1 + n_2 + 1}} + c_\alpha\right]^{-2}.$$

Beachtet man, daß $D^2(U)/\sigma_0^2$ sowie der Ausdruck innerhalb der eckigen Klammer der letzten Beziehung für $n_1 \to \infty$ über alle Grenzen wächst, so folgt

$$\lim_{n_1\to\infty} M = 0.$$

Somit ist im betrachteten Fall $p \geqq \frac{1}{2}$ dieser Test nicht konsistent, was durchaus auch unserer intuitiven Vorstellung entspricht.

Diese Ergebnisse über die Konsistenz des Tests von WILCOXON-MANN-WHITNEY stammen von VAN DANTZIG [1] (vgl. auch LEHMANN [2] und RÉNYI [10]). Die Konsistenz dieses Tests bezüglich der Menge der Verteilungsfunktionen, die (16.6.19) genügen, zeigten MANN und WHITNEY [1].

Das nachstehende, von GNEDENKO [11] angegebene Beispiel veranschaulicht die Inkonsistenz des Tests von WILCOXON-MANN-WHITNEY im Fall $p = \frac{1}{2}$.

Beispiel 16.6.2. Die Verteilungsfunktionen $F_1(x)$ und $F_2(x)$ seien stetig und mögen den Bedingungen

$$F_1(x) = \begin{cases} 0 & \text{für} \quad x \leqq b, \\ 1 & \text{für} \quad x \geqq c, \end{cases} \qquad F_2(x) = \begin{cases} 0 & \text{für} \quad x < a, \\ \frac{1}{2} & \text{für} \quad b \leqq x \leqq c, \\ 1 & \text{für} \quad x \geqq d \end{cases}$$

mit $a < b < c < d$ genügen. Wir haben hier

$$p = \int_{-\infty}^{\infty} F_2(x)\, dF_1(x) = \frac{1}{2}\big(F_1(c) - F_1(b)\big) = \frac{1}{2}.$$

Entnehmen wir jeder dieser Gesamtheiten Stichproben, so (wir benutzen hier die gleichen Bezeichnungen wie bisher) gilt

$$E(U) = \frac{1}{2} n_1 n_2, \qquad D^2(U) = \frac{1}{4} n_1 n_2.$$

Man erkennt sofort, daß für große n_1 und n_2 die Hälfte der beobachteten Werte von x_{2j} $(j = 1, 2, \ldots, n_2)$ im Intervall $[a, b]$ liegt, und für diese gilt somit $x_{2j} < x_{1l}$. Wir werden daher Werte der Stichprobenfunktion U beobachten, die nahe bei $\frac{1}{2} n_1 n_2$ liegen. Beim Überprüfen der Hypothese $H_0[F_1(x) \equiv F_2(x)]$ im Vergleich zur Alternativhypothese $H_1\left[F_1(x) \not\equiv F_2(x), p = \frac{1}{2}\right]$ mit Hilfe des Tests von WILCOXON-MANN-WHITNEY würde man zu der Annahme der falschen Hypothese H_0 sogar bei verhältnismäßig großen Werten von n_1 und n_2 geführt.

H. Der bisher betrachtete Test von WILCOXON-MANN-WHITNEY ist einseitig, denn die Alternativhypothese wird entweder durch (16.6.19) oder durch (16.6.20) gegeben. Man kann auch einen zweiseitigen Test von WILCOXON-MANN-WHITNEY betrachten, bei welchem die Nullhypothese nach wie vor die Hypothese $H_0[F_1(x) \equiv F_2(x)]$ bleibt, während die Alternativhypothese die Hypothese $H_1[F_1(x) \not\equiv F_2(x)]$ oder $H_1'\left(p \neq \frac{1}{2}\right)$ ist. Nun ist wie folgt zu verfahren. Für kleine n_1 und n_2 bestimmen wir zwei Zahlen u_1 und u_2 derart, daß die Stichprobenfunktion U von MANN-WHITNEY im Fall der Gültigkeit von H_0 den Beziehungen

$$P(U \leqq u_1) \leqq \frac{1}{2}\alpha, \quad P(U \leqq u_2) \leqq \frac{1}{2}\alpha$$

genügt, wobei α die Sicherheitswahrscheinlichkeit ist. Wir lehnen die Hypothese H_0 ab, wenn der beobachtete Wert u der Stichprobenfunktion U einer der beiden

Ungleichungen $u \leqq u_1$ oder $u \geqq u_2$ genügt. Für große n_1 und n_2 verwenden wir die Normalapproximation und lehnen die Hypothese H_0 ab, wenn der beobachtete Wert v der durch (16.6.22) definierten Stichprobenfunktion V der Beziehung $|v| \geqq v_{\alpha/2}$ genügt, wobei wir den Punkt $v_{\alpha/2}$ aus der Gleichung $\Phi(v_{\alpha/2}) = \frac{1}{2}\alpha$ bestimmen.

Unsere Betrachtungen zur Güte und Konsistenz des einseitigen Tests von MANN-WHITNEY lassen sich ohne Schwierigkeiten auf den zweiseitigen Test von MANN-WHITNEY übertragen. Insbesondere ist der Test von MANN-WHITNEY konsistent, wenn $H_0[F_1(x) \equiv F_2(x)]$ die Nullhypothese und $H_2\left[F_1(x) \not\equiv F_2(x),\ p \neq \frac{1}{2}\right]$ die Alternativhypothese ist.

16.7. Schlußbemerkungen

In den Arbeiten von WOLFOWITZ [4], HOEFFDING [1] und LEHMANN [3] findet der Leser viele Informationen über die Güte und Konsistenz des Tests von WALD-WOLFOWITZ und anderer nichtparametrischer Tests. Dem Problem der Güte des χ^2-Tests von PEARSON ist ebenfalls eine Arbeit von MANN und WALD [1] gewidmet. Die Arbeiten von BARTON, DAVID und MALLOWS [1] sowie von BARTON und DAVID [1] behandeln das Problem der Anwendbarkeit des Tests von WILCOXON-MANN-WHITNEY und des Tests von WALD-WOLFOWITZ zur Verifizierung des Tests über die Zufälligkeit einer Folge von Beobachtungen, die Elemente von zweierlei Art enthalten (vgl. Beispiel 14.2.3), sowie die Güte dieser Tests bei verschiedenen Alternativhypothesen.

16.8. Aufgaben und Ergänzungen

1. Der Parameter σ in einer Grundgesamtheit, in der das Merkmal X die Normalverteilung $N(0;\sigma)$ hat, ist unbekannt. Die Nullhypothese ist $H_0(\sigma = \sigma_0)$ und die Alternativhypothese $H_1(\sigma = \sigma_1)$. Man zeige, daß ein bester Test zur Überprüfung von H_0 im Vergleich zu H_1 anhand von Stichproben in diesem Fall die Menge w' ist, die durch die Ungleichungen

$$\bar{x}^2 + s^2 \leqq b \quad \text{für} \quad \sigma_0 > \sigma_1, \qquad \bar{x}^2 + s^2 \geqq b \quad \text{für} \quad \sigma_0 < \sigma_1$$

bestimmt ist; hierbei ermitteln wir die Konstante b aus der Bedingung $P(\xi \in w' | \sigma_0) = \alpha$. Man vergleiche hierzu das Beispiel 16.3.3.

2. Das Merkmal X habe die Dichtefunktion

$$f(x) = \begin{cases} \beta \exp[-\beta(x-\gamma)] & \text{für} \quad x - \gamma \geqq 0, \\ 0 & \text{für} \quad x - \gamma < 0. \end{cases}$$

Die Parameter β und γ sind unbekannt. Wir betrachten die Hypothesen $H_0(\beta = \beta_0, \gamma = \gamma_0)$ und $H_1(\beta = \beta_1, \gamma = \gamma_1)$, wobei die Menge der zulässigen Alternativhypothesen durch die Ungleichungen $\gamma_1 \leqq \gamma_0$ und $\beta_1 \geqq \beta_0$ bestimmt ist. Man zeige (NEYMAN und PEARSON [5]), daß ein bester Test zur Verifizierung der betrachteten Hypothesen der

kritische Bereich $w' = w'_1 + w'_2$ ist, wobei zu w'_1 die Punkte $x < \gamma_0$ und zu w'_2 die Punkte ξ gehören, die der Ungleichung

$$\bar{x} \leqq \frac{1}{\beta_1 - \beta_0}\left(\gamma_1\beta_1 - \gamma_0\beta_0 - \frac{1}{n}\log c + \log\frac{\beta_1}{\beta_0}\right) = A$$

genügen; dabei bestimmen wir die Konstante A aus der Sicherheitswahrscheinlichkeit α. Man beachte, daß dieser Test ein gleichmäßiger bester Test bezüglich β und γ ist. Man betrachte den Spezialfall, daß x die Exponentialverteilung mit $\gamma = 0$ hat.

3. Die Dichtefunktion $f(\xi, q)$ hängt von dem unbekannten Vektor $q = (Q_1, Q_2, \ldots, Q_r)$ ab. Wir bezeichnen mit q_0 bzw. q_1 die Vektoren $(Q_{10}, Q_{20}, \ldots, Q_{r0})$ bzw. $(Q_{11}, Q_{21}, \ldots, Q_{r1})$ und verifizieren die einfache Hypothese $H_0(q = q_0)$ im Vergleich zur einfachen Alternativhypothese $H_1(q = q_1)$ mit $q_1 \in \mathfrak{a}$.
Man zeige: Existiert ein gleichmäßig bester Text bezüglich der Klasse $\mathfrak{a}$ aller Alternativhypothesen, so läßt er sich auf der Grundlage der Schätzfunktion $q^*(\xi) = \big(Q_1^*(\xi), Q_2^*(\xi), \ldots, Q_r^*(\xi)\big)$ des Vektors q konstruieren, die man nach dem Maximum-Likelihood-Prinzip erhalten hat, wobei der Definitionsbereich des Vektors q die Menge $\mathfrak{a}$ ist.

4. (Wegen der Bezeichnungen siehe 16.3.D.) Es möge $f(\xi, Q)$, wobei Q eine beliebige reelle Zahl ist, die Dichtefunktion (wenn ξ vom stetigen Typ ist) oder die Wahrscheinlichkeitsfunktion (wenn ξ diskret ist) des Zufallspunktes ξ bezeichnen. Wir wollen annehmen, daß $f(\xi, Q)$ sich in der Form

$$f(\xi, Q) = C(Q)\exp[g(Q)\,T(\xi)]\,h(\xi) \tag{*}$$

darstellen läßt, wobei $g(Q)$ streng monoton wächst. Man zeige, daß zur Verifizierung der Hypothese $H_0(Q \leqq Q_0)$ im Vergleich zur Alternativhypothese $H_1(Q > Q_0)$ ein gleichmäßig bester verallgemeinerter Test existiert, der durch die Testfunktion

$$\varphi(\xi) = \begin{cases} 1 & \text{für } T(\xi) > c, \\ \gamma & \text{für } T(\xi) = c, \\ 0 & \text{für } T(\xi) < c \end{cases} \tag{**}$$

definiert ist; die Konstanten γ und c werden hierbei aus der Gleichung

$$P[T(\xi) > c \mid Q = Q_0] + \gamma P[T(\xi) = c \mid Q = Q_0] = \alpha$$

bestimmt. Falls die Funktion $g(Q)$ monoton fallend im strengen Sinne ist, während alle übrigen Bedingungen unverändert bleiben, sind in der Formel (**) die beiden Ungleichheitszeichen zu vertauschen.

5. Es möge x die Binomialverteilung gemäß (5.2.1) haben, und es sei p unbekannt. Man stelle die rechte Seite von (5.2.1) in der Form (*) gemäß der vorhergehenden Aufgabe dar und ermittle einen gleichmäßig besten Test zur Überprüfung der Hypothese $H_0(p \leqq p_0)$ im Vergleich zur Hypothese $H_1(p > p_0)$ mit der Sicherheitswahrscheinlichkeit α.

6. Es möge x die Normalverteilung $N(m; \sigma)$ haben.
   a) Wir wollen annehmen, daß m unbekannt und σ bekannt ist. Man verwende die Ergebnisse der Aufgabe 4, um einen gleichmäßig besten Test zur Überprüfung der Hypothese $H_0(m \leqq m_0)$ im Vergleich zur Hypothese $H_1(m > m_0)$ zu ermitteln. Man vergleiche den gewonnenen Test mit dem in 16.4 besprochenen Test.

b) Nun nehmen wir an, daß m bekannt und σ unbekannt ist. Man bestimme einen gleichmäßig besten Test zur Überprüfung der Hypothese $H_0(\sigma \leqq \sigma_0)$ im Vergleich mit der Hypothese $H_1(\sigma > \sigma_0)$. Man vergleiche hierzu Aufgabe 1.

7. Es möge $f(\xi, Q)$ die gleiche Bedeutung haben wie in Aufgabe 4 und sich in der Form (*) der genannten Aufgabe darstellen lassen. Zur Überprüfung der Hypothese $H_0(Q \leqq Q_1$ oder $Q \geqq Q_2)$ im Vergleich zur Hypothese $H_1(Q_1 < Q < Q_2)$ existiert ein gleichmäßig bester Test, der durch die Testfunktion

$$\varphi(\xi) = \begin{cases} 1 & \text{für } c_1 < T(\xi) < c_2, \\ \gamma_i & \text{für } T(\xi) = c_i \ (i = 1, 2), \\ 0 & \text{für } T(\xi) < c_1 \quad \text{oder} \quad T(\xi) > c_2 \end{cases}$$

gegeben ist; dabei werden die Konstanten $\gamma_1, \gamma_2, c_1, c_2$ aus der Gleichung

$$P[c_1 < T(\xi) < c_2 | Q = Q_i] + \gamma_1 P[T(\xi) = c_1 | Q = Q_i]$$

$$+ \gamma_2 P[T(\xi) = c_2 | Q = Q_i] = \alpha \qquad (i = 1, 2)$$

bestimmt.

8. Man verwende das Ergebnis der vorhergehenden Aufgabe, um einen gleichmäßig besten Test zur Überprüfung der Hypothese $H_0(m \leqq m_1$ oder $m \geqq m_2)$ im Vergleich zur Hypothese $H_1(m_1 < m < m_2)$ zu bestimmen, wobei m der unbekannte Mittelwert der Normalverteilung $N(m; 1)$ ist.

9. Man zeige: Ist $\varphi(\xi)$ die Testfunktion im Problem der Überprüfung einer Hypothese bezüglich Q und ist U eine erschöpfende Schätzfunktion des Parameters Q, so ist $E[\varphi(\xi) | U]$ ebenfalls eine Testfunktion, deren Gütefunktion gleich der Gütefunktion von $\varphi(\xi)$ ist.

10. Die Dichtefunktion $f(\xi, q)$ möge von dem unbekannten Vektor $q = (Q_1, Q_2, \ldots, Q_r, Q_{r+1})$ abhängen. Die zusammengesetzte Nullhypothese $H_0(Q_1 = Q_{10}, \ldots, Q_r = Q_{r0})$ möge den Wert des Parameters Q_{r+1} nicht spezifizieren. Wir setzen $q_0 = (Q_{10}, \ldots, Q_{r0}, Q_{r+1})$. Ferner mögen $\mathfrak{a}$ und $\mathfrak{A}$ die Menge der Vektoren q_0 und den Parameterraum darstellen. Wir setzen voraus, daß a) $f(\xi, q_0)$ hinreichend oft nach Q_{r+1} für nahezu alle Werte von Q_{r+1} differenzierbar ist und daß b) die Beziehung

$$\frac{\partial^2 \log f(\xi, q_0)}{\partial Q_{r+1}^2} = A + B \frac{\partial \log f(\xi, q_0)}{\partial Q_{r+1}}$$

gilt, wobei A und B zwar von Q_{r+1}, jedoch nicht von ξ abhängig sind.
Man zeige, daß die Menge w im Stichprobenraum für jede Sicherheitswahrscheinlichkeit α dann und nur dann ein dem ganzen Stichprobenraum ähnlicher Bereich bezüglich $\mathfrak{a}$ ist, wenn für $k = 1, 2, \ldots$ die Beziehung

$$\int_w \frac{\partial^k f(\xi, q_0)}{\partial Q_{r+1}^k} \, d\xi = 0$$

gilt.
Wir verwenden einen Test, der auf dem Maximum-Likelihood-Prinzip beruht: Die Dichtefunktion $f(\xi, q)$ möge von dem unbekannten Vektor $q = (Q_1, \ldots, Q_r, Q_{r+1}, \ldots, Q_{r+s})$ abhängen. Die Dichtefunktion $f(\xi, q)$ ist die Likelihood-Funktion (vgl. 13.7.A). $\mathfrak{A}$ möge die

Menge aller Vektoren q und $\mathfrak{a}$ die Menge aller Vektoren $q_0 = (Q_{10}, \ldots, Q_{r0}, Q_{r+1}, \ldots, Q_{r+s})$ bedeuten, wobei $Q_{10}, \ldots, Q_{r0}$ fest sind, während alle übrigen Q beliebig sind. Wir wollen die Hypothese $H_0(q \in \mathfrak{a})$ überprüfen. Hierzu setzen wir

$$\lambda = \frac{\sup\limits_{\mathfrak{a}} f(\xi, q)}{\sup\limits_{\mathfrak{A}} f(\xi, q)}. \tag{*}$$

Der auf dem Likelihood-Verhältnis beruhende Test lehnt die Hypothese H_0 ab, wenn $\lambda \leqq A$ ist, wobei A aus der Beziehung

$$\sup_{\mathfrak{a}} P(\lambda \leqq A \mid q \in \mathfrak{a}) = \alpha \tag{**}$$

bestimmt wird.

Wir bemerken, daß die Beziehung (**) schwächer ist als die Ähnlichkeit zum ganzen Stichprobenraum bezüglich der Menge $\mathfrak{a}$.

Für Zufallsvariable vom diskreten Typ läßt sich ein auf dem Likelihood-Verhältnis beruhender Test in analoger Weise konstruieren. Es ist dann allerdings die Dichtefunktion $f(\xi, q)$ durch die entsprechende Wahrscheinlichkeitsfunktion zu ersetzen.

11. Es mögen r normale Gesamtheiten $N(m_i; \sigma_i)$ $(i = 1, 2, \ldots, r)$ vorliegen, und es seien aus diesen Gesamtheiten r einfache Stichproben von jeweils dem Umfang n_i ausgewählt, wobei $n_1 + n_2 + \cdots + n_r = N$ gilt. Wir bezeichnen mit $\bar{x}_i$ bzw. mit s_i^2 den Erwartungswert bzw. die Varianz der i-ten Stichprobe. Die Nullhypothese $H_0(m_1 = m_2 = \cdots = m_r;\ \sigma_1 = \sigma_2 = \cdots = \sigma_r)$ spezifiziert weder den Erwartungswert noch die gemeinsame Varianz, während die Alternativhypothese H_1 sämtliche reellen m_i und positiven σ_i zuläßt. Man zeige, daß das Likelihood-Verhältnis die Form

$$\lambda = \prod_{i=1}^{r} \left(\frac{s_i^2}{s_0^2}\right)^{n_i/2}$$

mit $s_0^2 = \dfrac{1}{N} \sum\limits_{i=1}^{r} n_i[(\bar{x}_i - \bar{x}_0)^2 + s_i^2]$, wobei $\bar{x}_0 = \dfrac{1}{N} \sum\limits_{i=1}^{r} n_i \bar{x}_i$ ist.

12. (Fortsetzung). Wir behalten sämtliche Bezeichnungen und Voraussetzungen der vorhergehenden Aufgabe bei mit Ausnahme der Hypothese H_0, an deren Stelle wir die Hypothese $H_0'(\sigma_1 = \sigma_2 = \cdots = \sigma_r)$ ansetzen. Man zeige, daß

$$\lambda = \prod_{i=1}^{r} \left(\frac{s_i^2}{s_a^2}\right)^{n_i/2}$$

mit $s_a^2 = \dfrac{1}{N} \sum\limits_{i=1}^{r} n_i s_i^2$ gilt.

13. Man beweise: Sind die Parameter m und σ in einer normalen Grundgesamtheit unbekannt und wird σ von der Hypothese $H_0(m = m_0)$ nicht spezifiziert, so führt der auf dem Likelihood-Verhältnis beruhende Test auf den Studentschen t-Test.

14. Man beweise: Sind die Parameter m_1, σ_1 und m_2, σ_2 in den normalen Grundgesamtheiten $N(m_1; \sigma_1)$ und $N(m_2; \sigma_2)$ unbekannt, so gelten die folgenden Behauptungen:

a) Die Verifizierung der Hypothese $H_0(\sigma_1 = \sigma_2)$ (ohne irgendwelche Voraussetzungen über m_1 und m_2) führt bei Anwendung des Likelihood-Verhältnisses auf einen Test, der auf der Fisherschen Z-Verteilung beruht.

b) Die Verifizierung der Hypothese $H_0(m_1 = m_2)$ (wobei bekannt ist, daß $\sigma_1 = \sigma_2$ gilt) führt bei Anwendung des Likelihood-Verhältnisses auf den Studentschen t-Test.

15. Es möge x die Normalverteilung $N(m; \sigma)$ mit den Unbekannten m und σ haben. Wir betrachten die zusammengesetzte Hypothese $H_0(\sigma \leqq \sigma_0)$ im Vergleich zur Hypothese $H_1(\sigma > \sigma_0)$, wobei weder H_0 noch H_1 den Wert von m spezifiziert. Man zeige, daß der Test, der die Hypothese H_0 für $s^2 \geqq c$ ablehnt, wobei c aus der Gleichung $P[s^2 \geqq c \mid \sigma = \sigma_0] = \alpha$ bestimmt wird, ein gleichmäßig bester Test ist. Man vergleiche hierzu Aufgabe 6b.

16. Es möge $f(\xi, Q)$ die gleiche Bedeutung haben wie in Aufgabe 4 und sich in der Form (*) dieser Aufgabe darstellen lassen. Man zeige, daß zur Überprüfung der Hypothese $H_0(Q_1 \leqq Q \leqq Q_2)$ im Vergleich zur Hypothese $H_1(Q < Q_1$ oder $Q > Q_2)$ ein gleichmäßig bester unverfälschter Test existiert, der durch die Testfunktion

$$\varphi(\xi) = \begin{cases} 1 & \text{für} \quad T(\xi) < c_1 \quad \text{oder} \quad T(\xi) > c_2, \\ \gamma_i & \text{für} \quad T(\xi) = c_i \quad (i = 1, 2), \\ 0 & \text{für} \quad c_1 < T(\xi) < c_2 \end{cases} \tag{*}$$

gegeben ist; die Konstanten $c_1, c_2, \gamma_1, \gamma_2$ werden hierbei aus der Gleichung

$$E[\varphi(\xi) \mid Q = Q_i] = \alpha \qquad (i = 1, 2)$$

bestimmt.

17. Es möge $f(\xi, q)$ die gleichen Eigenschaften haben wie in der vorhergehenden Aufgabe Man zeige, daß zur Verifizierung der Hypothese $H_0(Q = Q_0)$ im Vergleich zur Hypothese $H_1(Q \neq Q_0)$ ein gleichmäßig bester, unverfälschter, durch die Testfunktion (*) von Aufgabe 16 gegebener Test existiert, wobei die Konstanten $c_1, c_2, \gamma_1, \gamma_2$ aus den Gleichungen

$$E[\varphi(\xi) \mid Q = Q_0] = \alpha, \qquad E[T(\xi)\, \varphi(\xi) \mid Q = Q_0] = \alpha E[T(\xi) \mid Q = Q_0]$$

bestimmt werden.

18. Es möge x die Normalverteilung $N(m; \sigma)$ mit unbekanntem m und σ haben.

a) Man zeige, daß für die Hypothese $H_0(\sigma \geqq \sigma_0)$ im Vergleich zur Hypothese $H_1(\sigma < \sigma_0)$, wobei beide Hypothesen H_0 und H_1 den Wert m nicht spezifizieren, der Test, der die Hypothese H_0 für $(s^2/\sigma_0^2) \leqq c$ ablehnt, wobei c aus der Gleichung $P(s^2 \leqq \sigma_0^2 c) = \alpha$ bestimmt wird, ein gleichmäßig bester unverfälschter Test ist (vgl. hierzu Aufgabe 15).

b) Man zeige, daß die gleichmäßig besten unverfälschten Tests zur Überprüfung der Hypothese $H_0(m \leqq m_0)$ im Vergleich zur Hypothese $H_1(m > m_0)$ oder der Hypothese $H_0(m = m_0)$ im Vergleich zur Hypothese $H_1(m \neq m_0)$ für den Fall, daß σ weder durch H_0 noch durch H_1 spezifiziert wird, auf dem Studentschen t-Test beruhen.

19. Es seien x_1 und x_2 unabhängige Zufallsvariable, beide mit der Poissonverteilung, und es möge x_i den Mittelwert λ_i $(i = 1, 2)$ haben. Wir wollen die zusammengesetzte Hypothese $H_0(\lambda_1 = \lambda_2)$ im Vergleich zur Hypothese $H_1(\lambda_1 \neq \lambda_2)$ überprüfen. Man leite einen gleich-

mäßig besten Test zur Überprüfung dieser Hypothese her (PRZYBOROWSKI und WILEŃSKI [1]).

20. (Bezeichnungen von 16.5). Der Test zur Überprüfung der einfachen Hypothese $H_0(Q = Q_0)$ im Vergleich zur Alternativhypothese $H_1(Q = Q_1)$ heißt *unverfälschter Test vom Typ A* mit der Sicherheitswahrscheinlichkeit α, wenn die Beziehungen (16.3.5) und (16.5.5) sowie die Beziehung

$$\left[\frac{\partial^2 M(w, Q)}{\partial Q^2}\right]_{Q=Q_0} = \max \tag{*}$$

erfüllt sind. Man zeige: Sind die Voraussetzungen des Satzes 16.5.1 erfüllt, existiert darüber hinaus die Ableitung $f_2(\xi, Q) = \dfrac{\partial^2 f(\xi, Q)}{\partial Q^2}$ in jedem inneren Punkt $Q \in K$ und gilt für jede Menge w im Stichprobenraum die Gleichung

$$\frac{\partial}{\partial Q} \int_w f_1(\xi, Q)\, d\xi = \int_w f_2(\xi, Q)\, d\xi,$$

so existiert unter den unverfälschten Tests vom Typ A ein bester Test w', zu dem die und nur die Punkte ξ gehören, für die

$$f_2(\xi, Q_0) \geqq c_1 f_1(\xi, Q_0) + c_2 f(\xi, Q_0) \tag{**}$$

gilt. Die Konstanten c_1 und c_2 bestimmen wir so, daß die Beziehungen (16.3.5), (16.5.5) und (*) erfüllt sind (NEYMAN und PEARSON [4]).

21. Man zeige, daß der beste unverfälschte Test im Beispiel 16.5.2 vom Typ A ist.

22. Das Merkmal X habe die Normalverteilung $N(0; \sigma)$. Wir wollen die Hypothese $H_0(\sigma = \sigma_0)$ im Vergleich zur Alternativhypothese $H_1(\sigma = \sigma_1 \neq \sigma_0)$ überprüfen. Man konstruiere hierzu einen besten unverfälschten Test vom Typ A.

23. Die Zufallsvariable X_i $(i = 1, 2)$ möge die stetige Verteilungsfunktion $F_i(x)$ haben. Wir bezeichnen mit x_{1l} $(l = 1, 2, \ldots, n_1)$ und x_{2j} $(j = 1, 2, \ldots, n_2)$ unabhängige Beobachtungen der Zufallsvariablen X_1 und X_2 und mit y_m $(m = 1, 2, \ldots, n_1 + n_2)$ die in aufsteigender Reihenfolge angeordneten Werte x_{1l} und x_{2j}; $m_1, m_2, \ldots, m_{n_2}$ seien die Ränge dieser y_m, die den x_{2j} entsprechen. Mit M_s $(s = 1, 2, \ldots, n_2)$ bezeichnen wir Zufallsvariable, die die Werte m_s durchlaufen. Man zeige (HOEFFDING [1], LEHMANN [3]): Gilt $F_2(x) = g[F_1(x)]$, wobei g eine stetige differenzierbare und nicht fallende Funktion und $g(0) = 0$ sowie $g(1) = 1$ ist, so gilt

$$P(M_1 = m_1, \ldots, M_{n_2} = m_{n_2}) = \frac{n_1!\, n_2!}{(n_1 + n_2)!} E\left[g'(u_{m_1}) \cdots g'\left(u_{m_{n_2}}\right)\right]; \tag{*}$$

dabei bedeutet u_{m_s} $(s = 1, 2, \ldots, n_2)$ den Wert der Positionsstichprobenfunktion $\zeta_{m_s}^{(n_1+n_2)}$ in einer einfachen Stichprobe aus einer im Intervall $[0, 1]$ gleichverteilten Grundgesamtheit.

24. (Fortsetzung). Man zeige (LEHMANN [3]): Ist $F_2(x) \equiv [F_1(x)]^k$ (k eine natürliche Zahl),

so hat die rechte Seite der Formel (*) die Gestalt

$$\frac{n_1!\,n_2!}{(n_1+n_2)!}\,k^{n_2}\prod_{s=1}^{n_2}\frac{\Gamma(m_s+ks-s)}{\Gamma(m_{s+1}+ks-s)}\,\frac{\Gamma(m_{s+1})}{\Gamma(m_s)}\,.$$

25. Unter einem *Rangtest* verstehen wir beim Problem zweier Stichproben einen Test, der lediglich von den beobachteten Rängen $m_1, m_2, \ldots, m_{n_2}$ abhängt.
Man konstruiere einen besten Rangtest zur Überprüfung von $H_0[F_1(x) \equiv F_2(x)]$ im Vergleich zur Alternativhypothese $H_1\big[F_2(x) \equiv g[F_1(x)]\big]$, wobei g die in der Aufgabe 23 genannten Eigenschaften hat (LEHMANN [3]).

26. (Fortsetzung). Man konstruiere einen besten Rangtest, um die Hypothese $H_0[F_1(x) \equiv F_2(x)]$ im Vergleich zur Alternativhypothese $H_1\big[F_2(x) \equiv (1-a)\,F(x) + a\big(F(x)\big)^2\big]$ zu prüfen, wobei $0 < a \leqq 1$ ist (LEHMANN [3], J. R. SAVAGE [1] und UHLMANN [1]).

17. ELEMENTE DER SEQUENTIALANALYSE

17.1. Einleitende Bemerkungen

A. Beim bisher besprochenen Testverfahren zur Überprüfung von Hypothesen war der Stichprobenumfang konstant. Bei einem derartigen Vorgehen nutzen wir gewöhnlich nicht die gesamte Information aus, die uns eine Stichprobe gibt. In der Tat, der Stichprobenumfang ist ja davon unabhängig, welche Elemente man als erste ausgewählt hat.

Leicht lassen sich Beispiele angeben, wo ein derartiges Verfahren offensichtlich ungünstig in dem Sinne ist, daß wir mehr Elemente als nötig auswählen. Betrachten wir zum Beispiel eine Gesamtheit, in der das Merkmal x normal $N(m; 1)$ verteilt ist, wobei der Mittelwert m unbekannt sei. Wir wollen die Hypothese $H_0(m = 0)$ testen. Zu diesem Zweck entnehmen wir der Gesamtheit eine Stichprobe vom Umfang n. Wir nehmen an, daß alle $n - 1$ zuerst gewählten Elemente in der Nähe des Wertes $x = 5$ liegen. Da die Standardabweichung 1 beträgt, ist es klar, daß die Wahl des n-ten Elements unsere Entscheidung über die Ablehnung der Hypothese H_0 nicht beeinflußt; denn die Wahrscheinlichkeit dafür, im Fall $m = 0$ nacheinander $n - 1$ Elemente in der Nähe von $x = 5$ zu erhalten, ist äußerst klein. Dieses Beispiel ist selbstverständlich nicht typisch, doch zeigt es äußerst klar den Nachteil eines Prüfverfahrens, in dem der Stichprobenumfang vorgegeben ist. Ganz anders liegen die Dinge bei einem sequentiellen Verfahren zur Prüfung einer Hypothese. Dieses Verfahren wollen wir im folgenden behandeln.

B. Ein *sequentielles Verfahren* zur Nachprüfung einer statistischen Hypothese nennt man ein Verfahren, bei dem der Stichprobenumfang nicht vorgegeben ist, sondern bei dem der Wert eines jeden neugewählten Elements darüber entscheidet, ob die zu verifizierende Hypothese anzunehmen oder abzulehnen ist oder ob die Untersuchung weiter fortgesetzt werden soll, d. h., ob ein weiteres Element in die Stichprobe gewählt werden soll. Ein bei einem sequentiellen Verfahren angewandter Test heißt *Sequentialtest*.

Die Problematik, die sich mit sequentiellen Prüfverfahren von statistischen Hypothesen befaßt, nennt man *Sequentialanalyse*.

Als Vorläufer der Begründung der Sequentialanalyse dürften wohl Dodge und Romig [1] anzusehen sein, die ein Schema einer zweistufigen Auswahl konstruiert hatten, bei dem von den Ergebnissen der ersten Stichprobe abhängig gemacht wird, ob noch eine zweite Stichprobe zu entnehmen ist. Bartky [1] betrachtete im

Spezialfall der Überprüfung einer Hypothese in bezug auf den Erwartungswert einer Binomialverteilung ein Schema, das dem Schema des sequentiellen Quotiententests ähnlich ist; dieser Test soll in 17.2 besprochen werden. Der Schöpfer der Theorie der Sequentialanalyse, auf den auch viele Anwendungen der Ergebnisse dieser Theorie zurückgehen, ist WALD. Die Grundideen der Sequentialanalyse sowie viele Anwendungen derselben findet der Leser in den Arbeiten von WALD [2], [4].

Ehe wir zu speziellen Betrachtungen übergehen, untersuchen wir die Frage, ob es möglich ist, daß die aufeinanderfolgenden Elementeuntersuchungen immer eine Fortsetzung des Verfahrens verursachen, aber keine Möglichkeit zur Entscheidung geben, ob die Hypothese abgelehnt oder angenommen werden soll. Solche Tests sind zwar theoretisch möglich; sie haben aber keine praktische Bedeutung. Die Sequentialtests, die wir betrachten wollen, haben die Eigenschaft, daß bei ihrer Anwendung die Wahrscheinlichkeit dafür, nach endlich vielen Beobachtungen eine Entscheidung treffen zu können, gleich 1 ist.

Einen Sequentialtest wollen wir (wie in der bisherigen Testtheorie) mit Hilfe der Gütefunktion oder auch mittels der Operationscharakteristik kennzeichnen. Außerdem wird er noch durch den *mittleren Stichprobenumfang* charakterisiert. Die Anzahl n der Beobachtungen, die für ein sequentielles Verfahren zur Entscheidung nötig sind, ist hier eine Zufallsvariable; diese Anzahl kann für jeden Versuch verschieden ausfallen. Der gegebene Sequentialtest wird durch den Durchschnittswert dieser Zufallsvariablen[1]) charakterisiert: Ein Test ist um so besser, je kleiner die durchschnittliche Beobachtungsanzahl bei seiner Anwendung ist.

17.2. Der sequentielle Quotiententest

In diesem Kapitel wollen wir die folgenden Bezeichnungen benutzen: Die Verteilung einer Zufallsvariablen x hänge vom Parameter Q ab; mit $f(x, Q)$ wollen wir die Dichte von Zufallsvariablen x bezeichnen, wenn x stetig ist, während für diskretes x mit $f(x, Q)$ die Wahrscheinlichkeitsfunktion der Zufallsvariablen x bezeichnet werden soll.

Wir beschreiben jetzt einen speziellen sequentiellen Test, nämlich den sequentiellen Quotiententest. Der Wert des Parameters Q in der Funktion $f(x, Q)$ sei unbekannt. Wir stellen die Hypothese $H_0(Q = Q_0)$ auf und wollen sie gegen die Alternativhypothese $H_1(Q = Q_1)$ testen; dabei sei Q_1 ein gewisser von Q_0 verschiedener Wert. Wie bisher sollen α bzw. β die Wahrscheinlichkeiten für einen Fehler erster bzw. zweiter Art bezeichnen. In Abhängigkeit von den Zahlen α und β bestimmen wir zwei Zahlen A und B, für die $0 < B < 1 < A$ ist. (Wie man diese Zahlen bestimmt, wollen wir später besprechen.)

[1]) Wir beschränken uns hier auf Fälle, in denen dieser Durchschnittswert endlich ist.

Wir entnehmen dieser Gesamtheit eine einfache Stichprobe. Wir stellen uns vor, daß wir eine Beobachtung besitzen, die mit x_1 bezeichnet sei. Weiter sei

$$p_{11} = f(x_1, Q_1), \qquad p_{01} = f(x_1, Q_0). \tag{17.2.1}$$

So ist also p_{11} im Fall einer diskreten Zufallsvariablen die Wahrscheinlichkeit dafür, daß $x = x_1$ ist, während p_{11} im Fall einer stetigen Zufallsvariablen der Wert der Dichte im Punkt $x = x_1$ ist, wenn die Hypothese H_1 richtig ist. Ähnliche Bedeutung hat p_{01}. Wir betrachten den Wert des Quotienten $\frac{p_{11}}{p_{01}}$.

Wenn dieser Quotient nun die Ungleichung

$$\frac{p_{11}}{p_{01}} \geqq A$$

erfüllt, lehnen wir die Hypothese H_0 zugunsten ihrer Alternativhypothese H_1 ab; wenn dagegen

$$\frac{p_{11}}{p_{01}} \leqq B$$

ist, nehmen wir die Hypothese H_0 an; ist schließlich

$$B < \frac{p_{11}}{p_{01}} < A,$$

so wählen wir in die Stichprobe noch ein Element, das mit x_2 bezeichnet sei.

Wir setzen weiter

$$p_{12} = f(x_1, Q_1)\, f(x_2, Q_1), \quad p_{02} = f(x_1, Q_0)\, f(x_2, Q_0)$$

und verfahren ähnlich wie oben, d. h., wir lehnen die Hypothese H_0 ab und nehmen die Hypothese H_1 an, wenn $\frac{p_{12}}{p_{02}} \geqq A$ ist; wir nehmen die Hypothese H_0 an, wenn $\frac{p_{12}}{p_{02}} \leqq B$ ist; schließlich nehmen wir eine dritte weitere Beobachtung, wenn die Ungleichungen

$$B < \frac{p_{12}}{p_{02}} < A$$

bestehen.

Falls eine Stichprobe, die aus $m - 1$ Elementen besteht, keine Entscheidung zu fällen erlaubt, entnehmen wir der Gesamtheit ein m-tes Element, und das

weitere Vorgehen hängt vom Wert des Quotienten

$$\frac{p_{1m}}{p_{0m}} = \frac{f(x_1, Q_1) f(x_2, Q_1) \cdots f(x_m, Q_1)}{f(x_1, Q_0) f(x_2, Q_0) \cdots f(x_m, Q_0)}$$

ab. Ist nämlich

$$\frac{p_{1m}}{p_{0m}} \geqq A,$$

so lehnen wir die Hypothese H_0 zugunsten der Hypothese H_1 ab, ist dagegen die Bedingung

$$\frac{p_{1m}}{p_{0m}} \leqq B$$

erfüllt, so nehmen wir die Hypothese H_0 an. Endlich wählen wir ein weiteres Element in die Stichprobe, wenn die Beziehung

$$B < \frac{p_{1m}}{p_{0m}} < A$$

besteht.

Dieses Prüfverfahren von Hypothesen nennt man den *sequentiellen Quotiententest.*

Das hier beschriebene Verfahren erfordert die Beantwortung der folgenden Fragen:

1. Wann ist die Wahrscheinlichkeit dafür, daß das Verfahren nach endlich vielen Schritten zu einer Entscheidung führt, gleich 1?

2. Wie bestimmt man die Operationscharakteristik dieses sequentiellen Quotiententests?

3. Wie findet man die mittlere Anzahl $E(n)$ der Beobachtungen, die zu einer Entscheidung nötig sind?

4. Wie bestimmt man A und B in Abhängigkeit von α und β? Wie soll man sich also vergewissern, daß die Wahrscheinlichkeiten für einen Fehler erster bzw. zweiter Art gleich α bzw. β im Fall einer stetigen Zufallsvariablen und nicht größer als α bzw. β im Fall einer diskreten Zufallsvariablen sind?

In den weiteren Darlegungen wollen wir uns bemühen, diese Fragen zu beantworten. Es ist bequemer, mit den Logarithmen der Zahlen A und B und den Größen

$$z_i = \log \frac{f(x_i, Q_1)}{f(x_i, Q_0)} \qquad (i = 1, 2, \ldots) \tag{17.2.2}$$

zu operieren, als mit den entsprechenden Zahlen selbst.

Das Prüfverfahren des betrachteten sequentiellen Quotiententests kann man

dann folgendermaßen beschreiben: Ist das m-te Element $(m \geqq 1)$ in die Stichprobe gewählt worden, so lehnen wir im Fall

$$\sum_{i=1}^{m} z_i \geqq \log A$$

die Hypothese H_0 zugunsten ihrer Alternative H_1 ab, während wir dann, wenn

$$\sum_{i=1}^{m} z_i \leqq \log B$$

gilt, die Hypothese H_0 annehmen. Sind dagegen die Ungleichungen

$$\log B < \sum_{i=1}^{m} z_i < \log A$$

erfüllt, so untersuchen wir das $(m+1)$-te Element.

Die Lösung der oben formulierten vier Probleme, die mit dem sequentiellen Quotiententest zusammenhängen, ergibt sich aus der Lösung derselben Probleme für ein allgemeineres Sequentialmodell, und zwar werden wir annehmen, daß die Zufallsvariablen z_i gewisse allgemeine Eigenschaften aufweisen, doch werden wir nicht voraussetzen, daß sie von der Form (17.2.2) sind.

17.3. Hilfssätze

A. Wir betrachten nun die Folge $\{z_i\}$ von unabhängigen Zufallsvariablen mit derselben Verteilung und endlicher Dispersion. Wir bezeichnen mit z eine Zufallsvariable, die dieselbe Verteilung wie die z_i hat, und mit Z_m die Summe der ersten m Glieder der Folge $\{z_i\}$:

$$Z_m = z_1 + z_2 + \cdots + z_m. \tag{17.3.1}$$

Ähnlich wie in 17.2 führen wir die Zahlen A und B ein, die die Ungleichungen $0 < B < 1 < A$ erfüllen. Dann gilt

$$\log B = b < 0, \qquad \log A = a > 0.$$

Mit n sei die kleinste natürliche Zahl bezeichnet, für die Z_n außerhalb des offenen Intervalls (b, a) liegt, für die also entweder $Z_n \leqq b$ oder $Z_n \geqq a$ ist.

Bestehen für jedes m $(m = 1, 2, \ldots)$ die Ungleichungen $b < Z_m < a$, so sagen wir, daß $n = \infty$ ist.

Satz 17.3.1. *Wenn die Dispersion der Zufallsvariablen z von Null verschieden ist, so ist die Wahrscheinlichkeit dafür, daß $n = \infty$ ist, gleich Null.*

Beweis. Es sei $c = a - b$. Wir teilen die unendliche Folge $z_1, z_2, \ldots$ in Abschnitte S_k zu r Gliedern ein, wobei r eine natürliche Zahl ist.

Der erste Abschnitt S_1 besteht aus den Gliedern $z_1, z_2, \ldots, z_r$, der zweite Abschnitt S_2 aus den Gliedern $z_{r+1}, z_{r+2}, \ldots, z_{2r}$ usw. Allgemein besteht der k-te Abschnitt S_k aus den Gliedern $z_{(k-1)r+1}, z_{(k-1)r+2}, \ldots, z_{kr}$.

Mit ζ_k wollen wir die Summe der Glieder des k-ten Abschnitts bezeichnen:

$$\zeta_k = \sum_{i=(k-1)r+1}^{kr} z_i.$$

Es sei $n = \infty$. Dann ist für jedes natürliche k die Ungleichung

$$\zeta_k^2 < c^2 \tag{17.3.2}$$

erfüllt.

In der Tat sind für jedes m nach Voraussetzung die Ungleichungen

$$b < \sum_{i=1}^{m} z_i < a$$

erfüllt. Für beliebige k und r bestehen also die Ungleichungen

$$b < \sum_{i=1}^{kr} z_i < a.$$

Daraus folgt

$$b - \sum_{i=1}^{(k-1)r} z_i < \zeta_k = \sum_{i=(k-1)r+1}^{kr} z_i < a < \sum_{i=1}^{(k-1)r} z_i.$$

Wenn wir noch

$$b < \sum_{i=1}^{(k-1)r} z_i < a$$

berücksichtigen, erhalten wir

$$-c = b - a < \zeta_k < a - b = c, \tag{17.3.3}$$

und hieraus folgt (17.3.2).

Um zu zeigen, daß die Wahrscheinlichkeit dafür, daß bei einem sequentiellen Quotiententest nur endlich viele Beobachtungen erforderlich sind, gleich 1 ist, genügt es, für ein geeignetes r zu zeigen, daß die Wahrscheinlichkeit dafür, daß die Ungleichung (17.3.2) für jedes k gilt, gleich 0 ist.

Wir setzen

$$\pi_k = P(\zeta_k^2 < c^2).$$

Da die Zufallsvariablen z_i $(i = 1, 2, \ldots)$ unabhängig sind und dieselbe Verteilung haben, sind die Zufallsvariablen ζ_k $(k = 1, 2, \ldots)$ gleichfalls unabhängig und gleichverteilt. Daraus folgt erstens, daß die Werte π_k für alle k dieselben sind (wir können also diese Wahrscheinlichkeit einfach mit π bezeichnen), und zweitens, daß die Wahrscheinlichkeit des Ereignisses, das daraus besteht, daß die Ungleichung (17.3.2) für $k = 1, 2, \ldots, j$ gilt, gleich π^j ist.

Um den Beweis zu beenden, genügt es zu zeigen, daß $\pi < 1$ ist.

Nach Voraussetzung ist die Dispersion $D^2(z)$ von 0 verschieden. Wenn wir die Unabhängigkeit der Zufallsvariablen z_i berücksichtigen, erhalten wir

$$E(\zeta_k^2) = E\left[\left(\sum_{i=(k-1)r+1}^{kr} z_i\right)^2\right] = r[E(z^2)] + r(r-1)[E(z)]^2 = r\left[D^2(z) + r\left(E(z)^2\right)\right].$$

Wählen wir also die Anzahl r der Glieder eines Abschnitts hinreichend groß, so wird auch $E(\zeta_k^2)$ beliebig groß; dann ist $\pi < 1$ (siehe Aufgabe 17.11.1).

Aus dem bewiesenen Satz 17.3.1 folgt somit: Besitzt die Zufallsvariable

$$z = \log \frac{f(x, Q_1)}{f(x, Q_0)}$$

eine endliche Varianz, so ist die Wahrscheinlichkeit dafür, daß das oben beschriebene Vorgehen beim sequentiellen Quotiententest zu einer Entscheidung für eine endliche Stichprobe führt, gleich 1.

Wir haben also die erste der vier Fragen beantwortet, die in 17.2 gestellt wurden.

B. Wir wollen jetzt gewisse zusätzliche Voraussetzungen über die Verteilung der Zufallsvariablen z machen und einen Satz beweisen, der für die weitere Darlegung der Sequentialanalyse wichtig ist.

Satz 17.3.2. *Es sei z eine Zufallsvariable, die die folgenden Bedingungen erfüllt:*

1. *Die Dispersion $D^2(z)$ existiert und ist von Null verschieden; es existiert also auch der Mittelwert $E(z)$.*
2. *Es existiert eine positive Zahl δ derart, daß die Ungleichungen*

$$P\left(z < \log(1-\delta)\right) > 0, \qquad P\left(z > \log(1+\delta)\right) > 0 \tag{17.3.4}$$

erfüllt sind.

3. *Für jede reelle Zahl h existiert der Mittelwert $E(e^{hz}) = g(h)$.*
4. *Es existieren die ersten beiden Ableitungen der Funktion $g(h)$, und man kann sie durch Differentiation unter dem Zeichen E erhalten; darunter verstehen wir eine Differentiation unter dem Integrationszeichen, wenn z eine stetige Zufallsvariable ist, und eine Differentiation unter dem Summationszeichen, wenn z*

eine diskrete Zufallsvariable ist. Dann gilt also

$$g'(h) = E(z e^{hz}), \tag{17.3.5}$$

$$g''(h) = E(z^2 e^{hz}). \tag{17.3.6}$$

Die Behauptung des Satzes lautet: *Wenn* $E(z) \neq 0$ *ist, gibt es einen einzigen reellen Wert* $h_0 \neq 0$, *der die Gleichung*

$$E(e^{h_0 z}) = 1 \tag{17.3.7}$$

erfüllt. Ist dagegen $E(z) = 0$, *so erfüllt nur der Wert* $h_0 = 0$ *die Gleichung* (17.3.7).

Beweis. Es sei $h > 0$ eine beliebige positive Zahl. Wir betrachten die Zufallsvariable $y = e^z$. Diese Zufallsvariable kann nur positive Werte annehmen, dasselbe gilt auch für die Zufallsvariable y^h. Nach Satz 3.3.1 erhalten wir daraus

$$g(h) = E(y^h) \geqq (1 + \delta)^h P\big(y^h \geqq (1 + \delta)^h\big) = (1 + \delta)^h P(e^z \geqq 1 + \delta). \tag{17.3.8}$$

Aus Bedingung 2 folgt $P(e^z \geqq 1 + \delta) > 0$. Wenn wir also die Beziehung

$$\lim_{h \to \infty} (1 + \delta)^h = \infty$$

beachten, so erhalten wir aus (17.3.8)

$$\lim_{h \to \infty} g(h) = \infty. \tag{17.3.9}$$

Ähnlich können wir zeigen, daß für ein beliebiges $h < 0$ die Beziehung

$$g(h) \geqq (1 - \delta)^h P(e^z \leqq 1 - \delta)$$

besteht. Aus Bedingung 2 und

$$\lim_{h \to -\infty} (1 - \delta)^h = \infty$$

ergibt sich

$$\lim_{h \to -\infty} g(h) = \infty. \tag{17.3.10}$$

Weiter betrachten wir die zweite Ableitung der Funktion $g(h)$, die im Sinne von Bedingung 4 durch die Formel (17.3.6) bestimmt ist. Wenn wir in dieser Formel die Bedingung 2 beachten, erhalten wir

$$g''(h) > 0. \tag{17.3.11}$$

Die Formeln (17.3.9) bis (17.3.11) zeigen, daß nur ein einziger reeller Wert h^*

existiert, für den die Funktion $g(h)$ ihr Minimum annimmt; dieser ist durch $g'(h^*) = 0$ bestimmt. Wenn $E(z) \neq 0$ ist, dann ist $g'(0) = E(z) \neq 0$, also $h^* \neq 0$ und $g(h^*) < g(0) = 1$.

Wie wir ferner erkennen, ist die Funktion $g(h)$ im Intervall $(-\infty, h^*)$ streng monoton abnehmend, während sie im Intervall (h^*, ∞) streng monoton wachsend ist. Da $g(0) = 1$ und $g(h^*) < 1$ ist, existiert also nur ein einziger Wert $h_0 \neq 0$ derart, daß $g(h_0) = E(e^{h_0 z}) = 1$. Ist dagegen $E(z) = g'(0) = 0$ und sind alle übrigen Bedingungen 1 bis 4 erfüllt, so erkennt man leicht aus dem Beweisverlauf, daß für diesen Fall $h = 0$ der einzige reelle Wert ist, für den $g(h) = 1$ ist. Der Satz 17.3.2 ist damit bewiesen.

17.4. Eine grundlegende Identität

Wir wollen jetzt eine grundlegende Identität beweisen, die eine große Rolle in der Sequentialanalyse spielt.

Es sei $\{z_i\}$ eine Folge von unabhängigen Zufallsvariablen, die dieselbe Verteilung wie die Zufallsvariable z haben. Mit Z_m sei die Summe der ersten m Glieder der Folge $\{z_i\}$ bezeichnet. Wenn die Zufallsvariable z die Bedingungen 1 *bis* 4 *aus* 17.3 *erfüllt, so gilt für jeden Punkt h aus dem Gebiet D, das durch die Ungleichung*

$$g(h) = E(e^{zh}) \geqq 1 \tag{17.4.1}$$

charakterisiert wird, die Identität

$$E\left\{e^{Z_n h}\big(g(h)\big)^{-n}\right\} = 1. \tag{17.4.2}$$

Dabei ist n wie in 17.3.A *definiert.*

Beweis. Wir betrachten die Gleichung

$$E(e^{Z_n h+(Z_N-Z_n)h}) = E(e^{Z_N h}) = E(e^{(z_1+z_2+\cdots+z_N)h}) = \big(g(h)\big)^N, \tag{17.4.3}$$

wobei N eine natürliche Zahl und n die kleinste natürliche Zahl ist, für die Z_n außerhalb des offenen Intervalls (b, a) liegt.

Wir bezeichnen mit P_N die Wahrscheinlichkeit dafür, daß $n \leqq N$ ist. Mit $E_N(u)$ wollen wir den bedingten Mittelwert der Zufallsvariablen u unter der Bedingung $n \leqq N$ und mit $E_N^*(u)$ den bedingten Mittelwert der Zufallsvariablen u unter der Bedingung $n > N$ bezeichnen.

Wenn wir diese Bezeichnungen benutzen, können wir (17.4.3) in der folgenden Form schreiben:

$$P_N E_N(e^{Z_n h+(Z_N-Z_n)h}) + (1-P_N)E_N^*(e^{Z_N h}) = \big(g(h)\big)^N. \tag{17.4.4}$$

Da für jedes feste $n \leqq N$ der Ausdruck $Z_N - Z_n = z_{n+1} + z_{n+2} + \cdots + z_N$ vom Wert $Z_n = z_1 + z_2 + \cdots + z_n$ unabhängig ist, können wir schreiben:

$$E_N(e^{Z_n h+(Z_N-Z_n)h}) = E_N(e^{Z_n h}\, e^{(Z_n-Z_n)h}) = E_N\left\{e^{Z_N h}\big(g(h)\big)^{N-n}\right\}. \quad (17.4.5$$

Aus den Formeln (17.4.4) und (17.4.5) erhalten wir die Identität in der Form

$$P_N E_N \{e^{Z_n h} (g(h))^{N-n}\} + (1 - P_N) E_N^* (e^{Z_N h}) = (g(h))^N .$$

Nach Division beider Seiten durch $(g(h))^N$ ergibt sich

$$P_N E_N \{e^{Z_n h} (g(h))^{-n}\} + (1 - P_N) \frac{E_N^* (e^{Z_N h})}{(g(h))^N} = 1 . \tag{17.4.6}$$

Ist nun die durch Formel (17.3.7) bestimmte Zahl h_0 negativ, so besteht das Gebiet D, das durch Formel (17.4.1) bestimmt ist, aus der ganzen reellen h-Achse mit Ausnahme des offenen Intervalls $(h_0, 0)$. Ist dagegen $h_0 > 0$, so besteht D aus der h-Achse ohne das offene Intervall $(0, h_0)$.

Da $\lim\limits_{N\to\infty} P(n \leqq N) = 1$ gilt, ist $\lim\limits_{N\to\infty} (1 - P_N) = 0$.

Wir behaupten, daß $E_N^* (e^{Z_N h})$ für $N \to \infty$ beschränkt ist. In der Tat ist dieser Ausdruck der Mittelwert von $e^{Z_N h}$ unter der Bedingung, daß $n > N$, also $b < Z_N < a$ ist. Es ist somit für $h > 0$

$$e^{bh} < e^{Z_N h} < e^{ah},$$

und für $h < 0$

$$e^{ah} < e^{Z_N h} < e^{bh}.$$

Daraus folgt, daß $E_N^* (e^{Z_N h})$ eine beschränkte Funktion ist.

Berücksichtigen wir schließlich noch, daß nur das Gebiet D derjenigen Werte h betrachtet wird, für welche $g(h) \geqq 1$ ist, so erhalten wir

$$\lim_{N\to\infty} \frac{(1 - P_N) E_N^* (e^{Z_N h})}{[g(h)]^N} = 0 \tag{17.4.7}$$

und

$$\lim_{N\to\infty} P_N E_N \{e^{Z_n h} [g(h)]^{-n}\} = E (e^{Z_N h} [g(h)]^{-n}\} . \tag{17.4.8}$$

Wenn wir (17.4.7) und (17.4.8) benutzen, so erhalten wir aus (17.4.6) die Formel (17.4.2).

Blom [1] bewies die Waldsche Grundidentität unter schwächeren Voraussetzungen, als wir sie angegeben haben, und verallgemeinerte darüber hinaus diese Identität für den Fall, daß die Zufallsvariablen z_i nicht die gleiche Verteilung haben. Blackwell und Girshick [1] verallgemeinerten die Grundidentität auf Zufallsvektoren.

17.5. Die Operationscharakteristik des sequentiellen Quotiententests

A. Die Operationscharakteristik des sequentiellen Quotiententests wollen wir hier kurz mit $L(Q)$ bezeichnen. Wir betrachten die Folge $\{z_i\}$ der unabhängigen Zufallsvariablen, die durch die Formel (17.2.2) bestimmt sind und den Bedingungen 1 bis 4 aus Satz 17.3.2 genügen. Die Zufallsvariable z sei durch

$$z = \log \frac{f(x, Q_1)}{f(x, Q_0)}$$

definiert. Wir setzen voraus, daß z die Bedingungen 1 bis 4 aus 17.3 erfüllt. Nach Satz 17.3.2 existiert für den Fall, daß $E(z) \neq 0$ gilt, ein einziger Wert $h_0 \neq 0$, der der Bedingung (17.3.7) genügt; ist dagegen $E(z) = 0$, so ist diese Bedingung nur für $h_0 = 0$ erfüllt.

Wir wollen uns vorläufig auf den Fall $E(z) \neq 0$ beschränken. Da die Verteilung der Zufallsvariablen z von Q abhängig ist, hängt auch der Wert h_0, der durch die Formel (17.3.7) bestimmt ist, von Q ab. Wir schreiben $h_0 = h_0(Q)$; aus (17.3.7) wird also

$$g(h_0(Q)) = E[e^{zh_0(Q)}] = 1. \tag{17.5.1}$$

Ist x eine stetige Zufallsvariable mit der Dichte $f(x, Q)$, so können wir diese Gleichung in der Gestalt

$$g(h_0(Q)) = \int_{-\infty}^{\infty} \left(\frac{f(x, Q_1)}{f(x, Q_0)}\right)^{h_0(Q)} f(x, Q)\, dx = 1 \tag{17.5.2}$$

schreiben.

Ist dagegen x eine diskrete Zufallsvariable, so nimmt die Gleichung (17.5.1) die Form

$$g(h_0(Q)) = \sum_j \left(\frac{f(x_j, Q_1)}{f(x_j, Q_0)}\right)^{h_0(Q)} f(x_j, Q) = 1 \tag{17.5.3}$$

an, wobei x_j $(j = 1, 2, \ldots)$ die Sprungstellen der Zufallsvariablen x sind, während mit $f(x_j, Q)$ in Übereinstimmung mit der in diesem Kapitel eingeführten Schreibweise die vom Parameter Q abhängige Wahrscheinlichkeit von $x = x_j$ bezeichnet ist.

Wir wollen jetzt die Identität (17.4.2) für den Wert $h = h_0$ aufschreiben. Da $g(h_0) = 1$ ist, können wir diese Identität in der Form

$$E_Q(e^{Z_n h_0(Q)}) = 1 \tag{17.5.4}$$

schreiben. Der Index Q soll darauf hinweisen, daß die linke Seite der Formel (17.5.4) eine Funktion des Parameters Q ist.

Ferner bezeichnen wir mit E_Q^* den bedingten Mittelwert der Zufallsvariablen $e^{Z_n h_0(Q)}$ unter der Bedingung, daß die Hypothese $H_0(Q = Q_0)$ angenommen wurde (d. h., daß $Z_n \leqq \log B$); analog bezeichnen wir mit E_Q^{**} den bedingten Mittelwert der Zufallsvariablen $e^{Z_n h_0(Q)}$ unter der Bedingung, daß die Hypothese $H_1(Q = Q_1)$ angenommen wurde (d. h., daß $Z_n \geqq \log A$). Unter Beachtung der Bedeutung des Symbols $L(Q)$ ergibt sich aus (17.5.4) die Formel

$$L(Q) E_Q^* + \big(1 - L(Q)\big) E_Q^{**} = 1$$

und daraus

$$L(Q) = \frac{E_Q^{**} - 1}{E_Q^{**} - E_Q^*}. \tag{17.5.5}$$

B. Wir wollen eine Näherungsformel für $L(Q)$ finden. Zu diesem Zweck betrachten wir statt der Ungleichungen $Z_n \leqq \log B$ und $Z_n \geqq \log A$ die Gleichungen $Z_n = \log B$ und $Z_n = \log A$. Setzen wir in den Ausdruck $e^{Z_n h_0(Q)}$ die Gleichung $Z_n = \log B$ ein, so erhalten wir

$$E_Q^* \left(e^{Z_n h_0(Q)}\right) \approx B^{h_0(Q)},$$

da wir bei $Z_n = \log B$ die Hypothese H_0 annehmen. Ähnlich ergibt sich, wenn man $Z_n = \log A$ einsetzt, die Näherungsformel

$$E_Q^{**} \left(e^{Z_n h_0(Q)}\right) \approx A^{h_0(Q)}.$$

Wenn wir diese Ausdrücke in (17.5.5) einsetzen, erhalten wir eine Näherungsformel für die Operationscharakteristik des sequentiellen Quotiententests:

$$L(Q) \approx \frac{A^{h_0(Q)} - 1}{A^{h_0(Q)} - B^{h_0(Q)}}. \tag{17.5.6}$$

Die Formel (17.5.6) wurde unter der Bedingung $E(z) \neq 0$, also für $h_0(Q) \neq 0$, hergeleitet. Wenn $E(z) = 0$ ist, dann ist auch $h_0(Q') = 0$, wobei Q' der Wert des Parameters Q ist, für den $E(z) = 0$ ist; die rechte Seite des Ausdrucks (17.5.6) hat dann im Punkt Q' den Grenzwert $\lim\limits_{Q \to Q'} L(Q) = L(Q')$, der gleich dem Quotienten der Ableitungen des Nenners und des Zählers nach Q im Punkt Q' ist. Wir erhalten also, wenn wir die Existenz der Ableitung von $h_0(Q)$ im Punkt Q' voraussetzen,

$$L(Q') \approx \frac{\log A}{\log A - \log B}. \tag{17.5.7}$$

Die Formel (17.5.2), eventuell auch (17.5.3), erlaubt es, Q als Funktion von h_0 zu bestimmen, während (17.5.6) den Wert von $L(Q)$ zu bestimmen erlaubt, wenn $h_0(Q)$ bekannt ist. Hieraus kann man für einen beliebigen reellen Wert

von h_0 Punkte der Ebene mit den Koordinaten $Q, L(Q)$ bestimmen. Die Kurve, die diese Punkte miteinander verbindet, ist eine angenäherte Darstellung der Operationscharakteristik.

Wir wollen uns überlegen, welcher Art die Näherung ist, die wir bei der Herleitung der Formeln (17.5.6) und (17.5.7) für die Operationscharakteristik anwandten.

Wenn wir $Z_n = \log A$ und $Z_n = \log B$ setzen, vernachlässigen wir die Möglichkeit, daß die Ungleichungen $Z_n > \log A$ oder $Z_n > \log B$ im Augenblick der Beendigung des sequentiellen Verfahrens gelten können. Mit anderen Worten, wir schreiben $Z_n = \log B$, obwohl in Wirklichkeit $Z_n < \log B$ ist, bzw. $Z_n = \log A$, obwohl in Wirklichkeit $Z_n > \log A$ ist.

Kann z nur die zwei Werte d und $-d$ annehmen, so kann Z_n nur gleich Kd sein, wobei K eine ganze Zahl, negativ oder positiv ist. Wenn zugleich $\frac{\log A}{d}$ und $\frac{\log B}{d}$ ganze Zahlen sind, so kann die Entscheidung über die Annahme der Hypothese nur bei $Z_n = \log B$ oder $Z_n = \log A$ fallen. In diesem Fall sind die Formeln (17.5.6) und (17.5.7) für $L(Q)$ vollkommen exakt. Erstreckt sich der Wertebereich der Zufallsvariablen z nicht nur auf die obenerwähnten Werte $-d$ und d, sind aber $|E_Q(z)|$ und $D_2^Q(z)$ nicht groß, so erweisen sich in der Praxis die Näherungsformeln für $L(Q)$ als gut.

Die Voraussetzung, daß Z_n im Augenblick der Entscheidung die Werte $\log A$ oder $\log B$ nicht übersteigt, wird häufig bei der Anwendung des sequentiellen Quotiententests gemacht.

17.6. Der Mittelwert $E(n)$

A. Wie schon in 17.1 gesagt wurde, charakterisiert die mittlere Anzahl der Beobachtungen $E(n)$ einen Sequentialtest. Für den sequentiellen Quotiententest, der zur Nachprüfung der Hypothese $H_0(Q = Q_0)$ im Vergleich mit der Alternativhypothese $H_1(Q = Q_1)$ dient, wollen wir die Größe $E(n)$ als Funktion des unbekannten Parameters Q finden. Dazu benutzen wir die grundlegende Identität (17.4.2).

Wir nehmen zuerst $E(z) \neq 0$ an. Wie Wald [3] zeigte, kann man, wenn die Bedingungen 1 bis 4 aus 17.3 erfüllt sind, die Identität beliebig oft unter dem Zeichen E nach h differenzieren. Wir differenzieren also (17.4.2) nach h im Punkt $h = 0$ und erhalten

$$E_Q\{Z_n e^{Z_n h}[g(h)]^{-n} - n e^{Z_n h}[g(h)]^{-n-1} g'(h)\}_{h=0} = 0. \tag{17.6.1}$$

Der untere Index Q soll wie vorher zeigen, daß dieser Mittelwert von Q abhängt. Da $g(0) = 1$ und $g'(0) = E(z)$ gilt, ergibt sich

$$E_Q[Z_n - n E_Q(z)] = 0,$$

und weil nach Voraussetzung $E_Q(z) \neq 0$ ist, erhalten wir

$$E_Q(n) = \frac{E_Q(Z_n)}{E_Q(z)}. \tag{17.6.2}$$

Wir wollen jetzt $E_Q(Z_n)$ bestimmen. Mit $E_Q^*(Z_n)$ bezeichnen wir den bedingten Mittelwert der Zufallsvariablen Z_n unter der Bedingung $Z_n \leqq \log B$ und mit $E_Q^{**}(Z_n)$ den bedingten Mittelwert der Zufallsvariablen Z_n unter der Bedingung $Z_n \geqq \log A$. Wir erhalten

$$E_Q(Z_n) = L(Q)E_Q^*(Z_n) + [1 - L(Q)]E_Q^{**}(Z_n). \tag{17.6.3}$$

Die Formel (17.6.2) nimmt dann folgende Gestalt an:

$$E_Q(n) = \frac{L(Q)E_Q^*(Z_n) + [1 - L(Q)]E_Q^{**}(Z_n)}{E_Q(z)}. \tag{17.6.4}$$

Wenn man die Differenzen $\log B - Z_n$ und $Z_n - \log A$ vernachlässigen kann, wenn man also annehmen kann, daß die Entscheidung bei $Z_n = \log B$ oder $Z_n = \log A$ fällt, so kann man für $E_Q^*(Z_n)$ die Größe $\log B$ und für $E_Q^{**}(Z_n)$ die Größe $\log A$ einsetzen. Die Formel für die mittlere Beobachtungsanzahl erhält also die Gestalt

$$E_Q(n) \approx \frac{L(Q)\log B + [1 - L(Q)]\log A}{E_Q(z)}. \tag{17.6.5}$$

Diese Formel ist exakt, wenn z nur die zwei Werte $-d$ und d annehmen kann, von denen in 17.5 die Rede war. Beschränkt sich der Wertebereich der Zufallsvariablen z nicht auf diese beiden Werte, sind aber $|E_Q(z)|$ und $D^2(z)$ nicht groß, so kann man, wie schon erwähnt, den Überschuß von Z_n über die Werte $\log A$ und $\log B$ außer acht lassen. Die Formel (17.5.6) ergibt dann eine Annäherung für den Mittelwert $E_Q(n)$.

B. Bisher setzten wir $E_Q(z) \neq 0$ voraus. Nun nehmen wir $E_{Q'}(z) = 0$ an. Nach zweimaligem Differenzieren der grundlegenden Identität nach h erhalten wir

$$E_{Q'}\left[\left\{\left[Z_n - n\frac{g'(h)}{g(h)}\right]^2 - \frac{ng''(h)g(h) - n\big(g'(h)\big)^2}{\big(g(h)\big)^2}\right\} e^{Z_n h}\big(g(h)\big)^{-n}\right] = 0. \tag{17.6.6}$$

Betrachtet man die Ableitung im Punkt $h = 0$ und berücksichtigt man, daß $g(0) = 1$, $g'(0) = E_{Q'}(z) = 0$ und $g''(0) = E_{Q''}(z^2) \neq 0$ ist, so erhält man aus (17.6.6) die Formel

$$E_{Q'}\big(Z_n^2 - nE_{Q'}(z^2)\big) = 0$$

und schließlich

$$E_{Q'}(n) = \frac{E_{Q'}(Z_n^2)}{E_{Q'}(z^2)}. \tag{17.6.7}$$

Wir wollen jetzt $E_{Q'}(Z_n^2)$ bestimmen. Mit $E_Q^*(Z_n^2)$ bzw. $E_Q^{**}(Z_n^2)$ wollen wir die bedingten Mittelwerte der Zufallsvariablen Z_n^2 unter der Bedingung $Z_n \leqq \log B$ bzw. $Z_n \geqq \log A$ bezeichnen. Wir haben dann

$$E_{Q'}(Z_n^2) = L(Q')E_{Q'}^*(Z_n^2) + [1 - L(Q')]E_{Q'}^{**}(Z_n^2). \tag{17.6.8}$$

Wenn man die Differenzen $\log B - Z_n$ und $Z_n - \log A$ vernachlässigen kann, nehmen $E_{Q'}^*(Z_n^2)$ und $E_{Q'}^{**}(Z_n^2)$ die Werte $(\log B)^2$ bzw. $(\log A)^2$ an.

Unter Beachtung der Formel (17.5.7) für $L(Q')$ im Fall, daß $E_{Q'}(z) = 0$ ist, ergibt sich

$$\begin{aligned} E_{Q'}(Z_n^2) &\approx \frac{\log A}{\log A - \log B}(\log B)^2 + \left[1 - \frac{\log A}{\log A - \log B}\right](\log A)^2 \\ &= -\log A \log B. \end{aligned} \tag{17.6.9}$$

Aus den Formeln (17.6.7) und (17.6.9) folgt schließlich

$$E_{Q'}(n) \approx -\frac{\log A \log B}{E_{Q'}(z^2)}. \tag{17.6.10}$$

Die letzte Formel hat unabhängig von Wald auch Wallis [1] hergeleitet.

Wir weisen den Leser darauf hin, daß die Formeln (17.6.2), (17.6.4) sowie die Näherungsformeln (17.6.5) und (17.6.10) für alle Zufallsvariablen gelten, die die Voraussetzungen des Satzes 17.3.2 erfüllen, und daß sie somit nicht nur für die in (17.2.2) definierten z_i gelten, die im sequentiellen Quotiententest auftreten. Blackwell [1], Blackwell und Girshick [1] sowie Wolfowitz [2] beschäftigten sich mit verschiedenen Verallgemeinerungen der Formel (17.6.2). In der zuletzt genannten Arbeit wird der Fall betrachtet, daß die Zufallsvariablen z_i abhängig sind und nicht die gleiche Verteilung haben (siehe auch Aufgabe 17.11.2).

17.7. Die Bestimmung der Zahlen *A* und *B*

Es bleibt nur noch eine der vier Fragen, die im Zusammenhang mit dem sequentiellen Quotiententest gestellt wurden, zu beantworten, nämlich die folgende: Wie bestimmt man die Zahlen A und B als Funktionen der Wahrscheinlichkeiten α bzw. β, mit welchen ein Fehler erster bzw. zweiter Art im sequentiellen Quotiententest gemacht werden kann? Wir wollen uns hier auf den Fall beschränken, daß man annehmen kann, die Entscheidung falle bei $Z_n = \log B$ oder $Z_n = \log A$.

Vor allem beachten wir, daß aus den Formeln (17.5.2) und (17.5.3) die Gleichungen

$$h_0(Q_0) = 1, \qquad h_0(Q_1) = -1 \tag{17.7.1}$$

folgen. Diese Ausdrücke setzen wir in (17.5.6) für $L(Q)$ ein. Unter Berücksichtigung von $L(Q_0) = 1 - \alpha$ und $L(Q_1) = \beta$ erhalten wir

$$L(Q_0) \approx \frac{A-1}{A-B} = 1 - \alpha, \tag{17.7.2}$$

$$L(Q_1) \approx \frac{\frac{1}{A} - 1}{\frac{1}{A} - \frac{1}{B}} = \beta. \tag{17.7.2'}$$

Die Auflösung dieses Gleichungssystems nach A und B ergibt

$$A \approx \frac{1-\beta}{\alpha}, \qquad B \approx \frac{\beta}{1-\alpha}. \tag{17.7.3}$$

Dies sind selbstverständlich nur Näherungsformeln, die aber für praktische Zwecke meist ausreichen.

17.8. Die Nachprüfung einer Hypothese über den Wert des Parameters p in einer Null-Eins-Verteilung

A. Wir befassen uns jetzt mit der Anwendung der hier dargelegten Theorie eines sequentiellen Quotiententests zur Nachprüfung einer Hypothese über den Wert des Parameters p in einer Null-Eins-Verteilung.

Die Zufallsvariable x habe eine Null-Eins-Verteilung, d. h., es sei $P(x = 1) = p$ und $P(x = 0) = 1 - p$ $(0 < p < 1)$, wobei p unbekannt sei. So bestehe z. B. eine Gesamtheit aus fehlerhaften und fehlerlosen Stücken. Dem Auslosen eines fehlerhaften Stücks ordnen wir die Zahl 1, dem Auslosen eines fehlerlosen die Zahl 0 zu. Gemäß der in diesem Abschnitt angenommenen Bezeichnungsweise ist

$$f(1, p) = p, \qquad f(0, p) = 1 - p. \tag{17.8.1}$$

Wir stellen die Hypothese $H_0(p = p_0)$ auf und wollen sie gegen die Alternativhypothese $H_1(p = p_1)$ testen. Dabei sind p_1 und p_0 bestimmte Werte, für die $0 < p_0 < p_1 < 1$ ist. Die Wahrscheinlichkeit, einen Fehler erster bzw. zweiter Art zu machen, sei gleich α bzw. β.

Wir wollen einen sequentiellen Quotiententest anwenden. Wiederum sei

$$z = \log \frac{f(x, p_1)}{f(x, p_0)}. \tag{17.8.2}$$

Die Zufallsvariable z kann zwei Werte annehmen, nämlich $z = \log p_1 - \log p_0$ für $x = 1$ und $z = \log(1 - p_1) - \log(1 - p_0)$ für $x = 0$. Zuerst wollen wir nachprüfen, ob die Zufallsvariable z die in 17.3 formulierten Bedingungen 1 bis 4 erfüllt.

Die Bedingung 1 ist erfüllt, da z eine endliche, von Null verschiedene Dispersion hat. Tatsächlich berechnen wir

$$E(z) = p \log \frac{p_1}{p_0} + (1 - p) \log \frac{1 - p_1}{1 - p_0} = \log \left[\left(\frac{p_1}{p_0}\right)^p \left(\frac{1 - p_1}{1 - p_0}\right)^{1-p}\right], \tag{17.8.3}$$

wobei p der wirkliche Wert des unbekannten Parameters ist. Weiter ist

$$E(z^2) = p \left(\log \frac{p_1}{p_0}\right)^2 + (1 - p) \left(\log \frac{1 - p_1}{1 - p_0}\right)^2.$$

Nach einer Umformung erhalten wir

$$D^2(z) = p(1 - p) \left[\log \frac{p_1(1 - p_0)}{p_0(1 - p_1)}\right]^2. \tag{17.8.4}$$

Da nach Voraussetzung $p_0 \neq p_1$ und p_0 sowie p_1 verschieden von Null und Eins sind, ist $D^2(z)$ endlich und von Null verschieden.

Wir gehen zur Bedingung 2 über. Es gilt

$$e^z = \frac{p_1}{p_0} \qquad \text{für} \quad x = 1, \tag{17.8.5}$$

$$e^z = \frac{1 - p_1}{1 - p_0} \qquad \text{für} \quad x = 0.$$

Nach Voraussetzung ist $p_1 > p_0$, also $1 - p_1 < 1 - p_0$. Wir setzen

$$\frac{p_1}{p_0} - 1 = \delta_1, \qquad 1 - \frac{1 - p_1}{1 - p_0} = \delta_2,$$

$$\delta = \frac{1}{2} \min(\delta_1, \delta_2).$$

Die Zahl δ ist also kleiner als jede der beiden Zahlen δ_1 und δ_2. Es bestehen daher die Beziehungen

$$P(e^z < 1 - \delta) = P\left(e^z = \frac{1 - p_1}{1 - p_0}\right) = 1 - p > 0, \tag{17.8.6}$$

$$P(e^z > 1 + \delta) = P\left(e^z = \frac{p_1}{p_0}\right) = p > 0.$$

Die Bedingung 2 ist also erfüllt.

Ferner gilt für jedes reelle h

$$g(h) = E(e^{hz}) = p\left(\frac{p_1}{p_0}\right)^h + (1 - p)\left(\frac{1 - p_1}{1 - p_0}\right)^h. \tag{17.8.7}$$

Aus den Voraussetzungen über p folgt, daß die Funktion $g(h)$ für jedes h existiert Die Bedingung 3 ist damit gleichfalls erfüllt.

Weiter gilt

$$g'(h) = p\left(\frac{p_1}{p_0}\right)^h \log\frac{p_1}{p_0} + (1 - p)\left(\frac{1 - p_1}{1 - p_0}\right)^h \log\frac{1 - p_1}{1 - p_0} = E(z e^{hz}). \tag{17.8.8}$$

Ähnlich kann man zeigen, daß

$$g''(h) = E(z^2 e^{hz}) \tag{17.8.9}$$

ist. Folglich ist auch die Bedingung 4 erfüllt.

B. Das sequentielle Verfahren hat hier folgenden Verlauf: Wir stellen uns vor, wir hätten schon m Beobachtungen angestellt. Dabei sei m_1-mal der Wert $x = 1$ und $(m - m_1)$-mal der Wert $x = 0$ beobachtet worden. Dann ist

$$Z_m = m_1 \log\frac{p_1}{p_0} + (m - m_1)\log\frac{1 - p_1}{1 - p_0}. \tag{17.8.10}$$

Die Konstanten A und B bestimmen wir aus (17.7.3).

Wir werden also die Hypothese H_0 ablehnen, wenn

$$m_1 \log\frac{p_1}{p_0} + (m - m_1)\log\frac{1 - p_1}{1 - p_0} \geqq \log A \tag{17.8.11}$$

ist, und die Hypothese H_0 annehmen, wenn

$$m_1 \log\frac{p_1}{p_0} + (m - m_1)\log\frac{1 - p_1}{1 - p_0} \leqq \log B \tag{17.8.12}$$

ist. Schließlich setzen wir die Untersuchung fort, wenn

$$\log B < m_1 \log \frac{p_1}{p_0} + (m - m_1) \log \frac{1 - p_1}{1 - p_0} < \log A \tag{17.8.13}$$

ist. Durch Umformung von (17.8.11) erhalten wir

$$m_1 \geqq \frac{\log A}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}} - m \frac{\log \frac{1 - p_1}{1 - p_0}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}}. \tag{17.8.11'}$$

Ähnlich erhalten wir aus den Formeln (17.8.12) und (17.8.13)

$$m_1 \leqq \frac{\log B}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}} - m \frac{\log \frac{1 - p_1}{1 - p_0}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}}, \tag{17.8.12'}$$

$$\begin{aligned} &\frac{\log B}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}} - m \frac{\log \frac{1 - p_1}{1 - p_0}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}} < m_1 \\ &< \frac{\log A}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}} - m \frac{\log \frac{1 - p_1}{1 - p_0}}{\log \frac{p_1}{p_0} - \log \frac{1 - p_1}{1 - p_0}}. \end{aligned} \tag{17.8.13'}$$

Die rechte Seite der Ungleichung (17.8.11′) wollen wir mit r_m bezeichnen und *Ablehnungszahl* nennen. Die linke Seite der Ungleichung (17.8.12′) wollen wir mit a_m bezeichnen; sie heiße *Annahmezahl.*

Dies bedeutet: Ist in einer einfachen Stichprobe vom Umfang m die Anzahl der fehlerhaften Stücke nicht kleiner als r_m, so lehnen wir die Hypothese $H_0(p = p_0)$ zugunsten ihrer Alternativhypothese $H_1(p = p_1)$ ab; ist dagegen die Anzahl der fehlerhaften Stücke nicht größer als a_m, so nehmen wir die Hypothese $H_0(p = p_0)$ an. Wir setzen die Untersuchung fort, falls die Anzahl der fehlerhaften Stücke größer als a_m und kleiner als r_m ist.

Ehe wir an die Ausführung der sequentiellen Versuche schreiten, können wir für gegebene p_0, p_1, α und β die Zahlen a_m und r_m berechnen.

Die Größen r_m und a_m sind lineare Funktionen des Stichprobenumfangs m. Man

kann sie also in der Form

$$r_m = cm + d_1 \quad \text{bzw.} \quad a_m = cm + d_2$$

schreiben.

Leicht bestätigt man, daß für $p_1 > p_0$ die Ungleichungen $c > 0$, $d_1 > 0$ und $d_2 < 0$ bestehen.

Der typische Verlauf dieser Funktionen in der m, m_1-Ebene ist in Abb. 17.8.1 dargestellt, ohne daß die Parameterwerte c, d_1 und d_2 präzisiert wurden. Wie man aus dieser Zeichnung ersieht, entspricht das oberhalb der Geraden r_m gelegene

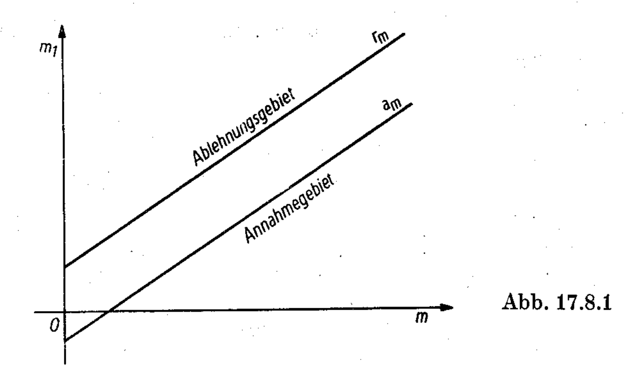


Abb. 17.8.1

Gebiet einschließlich der Geraden r_m den Testergebnissen (m, m_1), für welche wir die Hypothese H_0 ablehnen (wir nennen deshalb dieses Gebiet *Ablehnungsgebiet*). Ähnlich nennen wir das unterhalb der Geraden a_m gelegene Gebiet mit Einschluß der Geraden a_m das *Annahmegebiet*.

Liegt ein Punkt (m, m_1) zwischen den Geraden r_m und a_m, so setzen wir die Probenentnahme fort.

Wir wollen jetzt die Operationscharakteristik dieses Beispiels finden. Wenn wir diejenigen Werte von p auslassen, für die $E(z) = 0$ gilt, erhalten wir für die übrigen Werte von p wegen der Formel (17.5.3)

$$p\left(\frac{p_1}{p_0}\right)^{h_0(p)} + (1-p)\left(\frac{1-p_1}{1-p_0}\right)^{h_0(p)} = 1, \tag{17.8.14}$$

wobei $h_0(p) \neq 0$ ist.

Aus dieser Gleichung läßt sich p als Funktion von h_0 bestimmen:

$$p = \frac{1 - \left(\dfrac{1-p_1}{1-p_0}\right)^{h_0}}{\left(\dfrac{p_1}{p_0}\right)^{h_0} - \left(\dfrac{1-p_1}{1-p_0}\right)^{h_0}}. \tag{17.8.15}$$

Weiter erhalten wir aus der Formel (17.5.6) für $L(p)$, wenn wir die Werte A und B nach den Formeln (17.7.3) bestimmen,

$$L(p) = \frac{\left(\frac{1-\beta}{\alpha}\right)^{h_0} - 1}{\left(\frac{1-\beta}{\alpha}\right)^{h_0} - \left(\frac{\beta}{1-\alpha}\right)^{h_0}}. \tag{17.8.16}$$

Die Formeln (17.8.15) und (17.8.16) stellen eine Parametergleichung der Kurve $L(p)$, also der Operationscharakteristik dar.

Die mittlere Beobachtungsanzahl $E(n)$ berechnet man aus (17.6.5), indem man (17.8.15), (17.8.16), die Formel (17.7.3) für die Konstanten A und B und die Formel (17.8.3) für $E(z)$ benutzt.

Der Leser findet leicht den Wert von $L(p)$ für den Wert p, für welchen $E(z) = 0$ ist, und die mittlere Beobachtungsanzahl für diesen Wert p.

Wir wollen für das hier betrachtete Problem die folgenden Zahlenwerte annehmen:

$$p_0 = 0{,}05, \qquad p_1 = 0{,}1, \qquad \alpha = \beta = 0{,}05;$$

p sei der unbekannte Ausschußanteil in einer Gesamtheit, die fehlerhafte und fehlerlose Stücke enthält. Wir erhalten hier

$$\log A = \log \frac{1-\beta}{\alpha} = \log 19 = 1{,}27875,$$

$$\log B = \log \frac{\beta}{1-\alpha} = -\log 19 = -1{,}27875.$$

Die Zufallsvariable z, die durch (17.8.2) definiert ist, kann zwei Werte annehmen, nämlich den Wert $z = \log 2 = 0{,}30103$ für den Fall $x = 1$ (d. h. für den Fall, daß in der Stichprobe ein fehlerhaftes Stück vorkommt) und $z = \log \frac{18}{19}$ $= -0{,}02348$ für $x = 0$ (wenn ein fehlerloses Stück ausgelost wurde). Der sequentielle Quotiententest besteht hier darin, daß wir die Hypothese $H_0(p = 0{,}05)$ zugunsten der Hypothese $H_1(p = 0{,}1)$ ablehnen, wenn es sich nach dem Auslosen des m-ten Stückes $(m = 1, 2, \ldots)$ zeigt, daß sich in dieser Stichprobe m_1 fehlerhafte und $m - m_1$ fehlerlose Stücke befinden, wobei unter Berücksichtigung der Formel (17.8.11) für Z_m die Ungleichung

$$0{,}30103\, m_1 - 0{,}02348\,(m - m_1) \geqq 1{,}27875$$

erfüllt ist.

Ist dagegen

$$0{,}30103\, m_1 - 0{,}02348\, (m - m_1) \leqq -1{,}27875,$$

so nehmen wir die Hypothese $H_0(p = 0{,}05)$ an, und bestehen schließlich die Ungleichungen

$$-1{,}27875 < 0{,}30103\, m_1 - 0{,}02348\, (m - m_1) < 1{,}27875,$$

so setzen wir die Untersuchung fort, d. h., wir wählen ein weiteres $(m + 1)$-tes Stück in die Stichprobe.

Die Ablehnungszahlen r_m der Hypothese H_0, die durch die rechte Seite der Ungleichung (17.8.11') bestimmt sind, nehmen hier die Gestalt

$$r_m = \frac{1{,}27875}{0{,}30103 + 0{,}02348} - m\, \frac{-0{,}02348}{0{,}30103 + 0{,}02348} = 3{,}94 + 0{,}072\, m$$

an. Ähnlich finden wir

$$a_m = \frac{-1{,}27875}{0{,}30103 + 0{,}02348} - m\, \frac{-0{,}02348}{0{,}30103 + 0{,}02348} \doteq -3{,}94 + 0{,}072\, m.$$

Die Operationscharakteristik dieses Tests können wir aus den Formeln (17.8.15) und (17.8.16) bestimmen. Sie ist durch die folgenden Parametergleichungen gegeben:

$$p = \frac{1 - \left(\frac{18}{19}\right)^{h_0}}{2^{h_0} - \left(\frac{18}{19}\right)^{h_0}}, \qquad L(p) = \frac{19^{h_0} - 1}{19^{h_0} - \left(\frac{1}{19}\right)^{h_0}}.$$

Wie bekannt, ist nach (17.7.2) und (17.7.2')

$$L(0{,}05) = 0{,}95, \qquad L(0{,}10) = 0{,}05.$$

Wir berechnen noch die mittlere Beobachtungsanzahl, wenn die Hypothese $H_0(p = 0{,}05)$ bzw. die Hypothese $H_1(p = 0{,}1)$ richtig ist. Aus Formel (17.6.5) erhalten wir

$$E_{0,05}(n) = \frac{L(0{,}05)(-1{,}27875) + [1 - L(0{,}05)]\, 1{,}27875}{E_{0,05}(z)},$$

aus (17.8.3) ergibt sich

$$E_{0,05}(z) = \log 2^{0,05} + \log \left(\frac{18}{19}\right)^{0,95} = -0{,}00725.$$

Schließlich erhalten wir

$$E_{0,05}(n) \approx 159.$$

Auf ähnliche Weise ergibt sich

$$E_{0,1}(n) \approx 128.$$

Es könnte scheinen, daß die hier erhaltenen mittleren Beobachtungsanzahlen sehr groß sind. Wir zeigen jedoch, daß die Anwendung eines Sequentialtests sich günstiger als die Anwendung eines klassischen Tests auswirkt, der nach der Methode von NEYMAN-PEARSON (Kap. 16) gebildet wird.

Zu diesem Zweck wollen wir die betrachtete Hypothese $H_0(p = 0{,}05)$ gegen die Alternative $H_1(p = 0{,}1)$ testen, und zwar durch die Anwendung der klassischen Methode für

$$n = \frac{1}{2}\,[E_{0,05}(n) + E_{0,10}(n)] \approx 144.$$

Dabei wird weiter angenommen, daß $\alpha = 0{,}05$ ist. Zunächst wollen wir einen Untersuchungsplan aufstellen. Wir müssen also eine Zahl k finden, so daß wir die Hypothese $H_0(p = p_0 = 0{,}05)$ ablehnen, wenn die Anzahl der Ausschußstücke größer als k ist; dabei soll die Wahrscheinlichkeit, einen Fehler erster Art zu machen, nicht größer als 0,05 sein. Diese Zahl k können wir aus der Relation

$$\sum_{j=k+1}^{144} \binom{144}{j} 0{,}05^j \cdot 0{,}95^{144-j} \leqq 0{,}05$$

bestimmen. Da n groß ist, benutzen wir zur Bestimmung von k die Poissonsche Verteilung mit dem Mittelwert $np_0 = 7{,}2$ (siehe 5.5). Wir erhalten $k = 12$.

Jetzt berechnen wir die Wahrscheinlichkeit β für einen Fehler zweiter Art. Wir setzen also voraus, daß $p = p_1 = 0{,}1$ ist, und berechnen den Wert

$$\beta = 1 - \sum_{j=13}^{144} \binom{144}{j} 0{,}1^j \cdot 0{,}9^{144-j}.$$

Diese Größe finden wir aus der Poissonschen Verteilung mit dem Mittelwert $np_1 = 14{,}4$. Wir erhalten hier $\beta = 0{,}3203$. Die Wahrscheinlichkeit für einen Fehler zweiter Art ist hier also mehr als sechsmal so groß wie bei Anwendungen eines sequentiellen Quotiententestes mit ungefähr derselben durchschnittlichen Beobachtungsanzahl.

Um die Frage nach den Vorteilen vollständig zu beantworten, die wir für dieses Problem durch Anwendung eines sequentiellen Quotiententests statt einer klassischen Methode mit konstantem Stichprobenumfang erreichen, berechnen

wir den Wert n, der bei Anwendung der klassischen Methode auf das betrachtete Zahlenbeispiel nötig ist, um α und β nahe an 0,05 zu erreichen.

Da n hier bestimmt ziemlich groß sein wird (größer als 144), können wir den Moivre-Laplaceschen Grenzwertsatz benutzen. Wir bezeichnen mit y die Anzahl der Ausschußstücke in einer Stichprobe vom Umfang n. Diese Zufallsvariable ist asymptotisch normal $N\left(0{,}05\,n;\ \sqrt{0{,}0475\,n}\right)$ verteilt, wenn die Hypothese $H_0(p = 0{,}05)$ erfüllt ist, und asymptotisch normal $N\left(0{,}1\,n;\ \sqrt{0{,}09\,n}\right)$ verteilt, wenn die Hypothese $H_1(p = 0{,}1)$ erfüllt ist.

Um diese Hypothesen zu testen, wählen wir als kritische Region w die Menge der Punkte ξ im Stichprobenraum, für welche die Gleichung

$$P(\bar{y} \geqq A) = 0{,}05$$

erfüllt ist, wobei $\frac{y}{n} = \bar{y}$ normal $N\left(0{,}05;\ \sqrt{\frac{0{,}0475}{n}}\right)$ verteilt ist. Die unbekannten Werte A und n bestimmen wir aus dem System der zwei Gleichungen, die sich daraus ergeben, daß $\alpha = \beta = 0{,}05$ ist:

$$\frac{(A - 0{,}05)\sqrt{n}}{\sqrt{0{,}0475}} = 1{,}65, \qquad \frac{(A - 0{,}1)\sqrt{n}}{\sqrt{0{,}09}} = -1{,}65.$$

Aus diesem Gleichungssystem erhalten wir $n \approx 292$, $A \approx 0{,}071$.

Da A der kritische Wert von $\bar{y}$ ist, beträgt der kritische Wert von y dann $An = 20{,}7 \approx 21$; wir lehnen also die Hypothese $H_0(p = 0{,}05)$ zugunsten der Hypothese $H_1(p = 0{,}1)$ ab, wenn wir in einer einfachen, 292 Stücke zählenden Stichprobe 21 oder mehr Ausschußstücke vorfinden. Die Wahrscheinlichkeit für einen Fehler erster Art wie auch die Wahrscheinlichkeit für einen Fehler zweiter Art werden wegen der diskreten Verteilung von y und der benutzten normalen Annäherung ungefähr gleich 0,05 sein.

Wir kommen also zu dem Schluß, daß man eine Stichprobe mit $n = 292$ Elementen erheben muß, um einen gleichguten Test zu erhalten, also beinahe zweimal soviel, wie der durchschnittliche Stichprobenumfang bei Anwendung eines sequentiellen Quotiententests beträgt. Die Anwendung von Sequentialtests ist also zweifellos vorteilhafter.

17.9. Die Nachprüfung einer Hypothese über den Mittelwert Q einer Normalverteilung

Es sei x normal $N(Q; 1)$ verteilt, wobei Q den unbekannten Mittelwert bedeutet. Wir stellen die Hypothese $H_0\ (Q = Q_0)$ und die Alternativhypothese $H_1\ (Q = Q_1)$ auf, wobei $Q_1 > Q_0$ ist. Mit den bisherigen

Bezeichnungen ist

$$f(x, Q_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-Q_0)^2}, \tag{17.9.1}$$

$$f(x, Q_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-Q_1)^2}, \tag{17.9.2}$$

$$z = \log \frac{f(x, Q_1)}{f(x, Q_0)} = \frac{1}{2} [(x - Q_0)^2 - (x - Q_1)^2]$$

$$= (Q_1 - Q_0)x + \frac{1}{2} (Q_0^2 - Q_1^2). \tag{17.9.3}$$

Wie der Leser mit Leichtigkeit nachprüft, erfüllt die Zufallsvariable z die in 17.3 formulierten Bedingungen 1 bis 4. Wir erhalten

$$Z_m = \sum_{i=1}^{m} z_i = (Q_1 - Q_0) \sum_{i=1}^{m} x_i + \frac{m}{2} (Q_0^2 - Q_1^2), \tag{17.9.4}$$

wobei die x_i $(i = 1, 2, \ldots, m)$ unabhängige Beobachtungen der Zufallsvariablen x sind.

Wir bestimmen nun die Operationscharakteristik dieses Tests. Aus der Formel (17.5.2) erhalten wir

$$g\big(h_0(Q)\big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-Q)^2}{2}\right) \left[\frac{\exp\left(-\frac{(x-Q_1)^2}{2}\right)}{\exp\left(-\frac{(x-Q_0)^2}{2}\right)}\right]^{h_0(Q)} dx = 1 \tag{17.9.5}$$

Nach Umformung ergibt sich

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2 - 2\big(Q + h_0(Q)(Q_1 - Q_0)\big)x}{2}\right.$$

$$\left. - \frac{Q^2 + h_0(Q)(Q_1^2 - Q_0^2)}{2}\right\} dx = 1. \tag{17.9.6}$$

Den Exponenten im letzten Integral schreiben wir in der Form

$$-\frac{1}{2}\{x-(Q+h_0(Q)(Q_1-Q_0))\}^2$$

$$-\frac{1}{2}h_0(Q)(Q_1-Q_0)\big(Q_1+Q_0-2Q-h_0(Q)(Q_1-Q_0)\big).$$

Da

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\left(\frac{x-(Q+h_0(Q)(Q_1-Q_0))}{2}\right)^2\right\}dx=1 \tag{17.9.7}$$

ist, muß die Gleichung

$$h_0(Q)=\frac{Q_1+Q_0-2Q}{Q_1-Q_0} \tag{17.9.8}$$

gelten, damit (17.9.6) erfüllt ist. Daraus folgt, daß das durch Formel (17.9.8) bestimmte $h_0(Q)$ eine Lösung der Gleichung (17.9.5) ist. Da die Zufallsvariable z die Bedingungen 1 bis 4 aus 17.3 erfüllt, gibt es für jeden Wert von Q nur einen einzigen Wert $h_0(Q)$, der (17.9.5) erfüllt. Daher ist das durch (17.9.8) gegebene $h_0(Q)$ die einzige Lösung.

Die Operationscharakteristik erhalten wir aus (17.5.6) und (17.9.8). Somit ist

$$L(Q)=\frac{\left(\frac{1-\beta}{\alpha}\right)^{h_0(Q)}-1}{\left(\frac{1-\beta}{\alpha}\right)^{h_0(Q)}-\left(\frac{\beta}{1-\alpha}\right)^{h_0(Q)}}. \tag{17.9.9}$$

Diese Formel gilt für alle Werte Q mit $h_0(Q)\neq 0$.

Leicht erkennt man, daß $h_0(Q)=0$ für

$$Q'=\frac{Q_1+Q_0}{2}$$

ist. Also besteht nach Formel (17.5.7) für diesen Wert Q' die Gleichung

$$L(Q')=\frac{\log\frac{1-\beta}{\alpha}}{\log\frac{1-\beta}{\alpha}-\log\frac{\beta}{1-\alpha}}. \tag{17.9.10}$$

Wir berechnen jetzt die mittlere Beobachtungsanzahl $E_Q(n)$ als Funktion des Parameters Q.

Für $Q \neq Q'$ haben wir nach Formel (17.6.5)

$$E_Q(n) = \frac{L(Q)\log B + [1 - L(Q)]\log A}{E_Q(z)}. \tag{17.9.11}$$

Die Operationscharakteristik $L(Q)$ ist durch (17.9.9) gegeben. Wir müssen noch $E_Q(z)$ bestimmen. Aus (17.9.3) ergibt sich unter Berücksichtigung der Beziehung $E(x) = Q$ die Gleichung

$$E_Q(z) = (Q_1 - Q_0)\,Q + \frac{1}{2}\,(Q_0^2 - Q_1^2). \tag{17.9.12}$$

Im Punkt Q' gilt die Beziehung $E_{Q'}(z) = 0$; man kann also hier die Formel (17.9.11) nicht anwenden. Die durchschnittliche Beobachtungsanzahl finden wir für diesen Fall aus

$$E_{Q'}(n) = -\frac{\log\dfrac{1-\beta}{\alpha}\log\dfrac{\beta}{1-\alpha}}{E_{Q'}(z^2)}. \tag{17.9.13}$$

Es gilt

$$E_{Q'}(z^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{4}\,[2(Q_1 - Q_0)\,x + (Q_0^2 - Q_1^2)]^2$$

$$\times \exp\left\{-\frac{\left[x - \dfrac{Q_1 + Q_0}{2}\right]^2}{2}\right\}dx.$$

Nach einfachen Umformungen erhalten wir

$$E_{Q'}(z^2) = (Q_0 - Q_1)^2. \tag{17.9.14}$$

Schließlich ergibt sich

$$E_{Q'}(n) = -\frac{\log\dfrac{1-\beta}{\alpha}\log\dfrac{\beta}{1-\alpha}}{(Q_0 - Q_1)^2}. \tag{17.9.15}$$

Wir wollen speziell die Werte $E_{Q_0}(n)$ und $E_{Q_1}(n)$ berechnen. Da

$$L(Q_0) = 1 - \alpha,$$

$$L(Q_1) = \beta$$

gilt, erhalten wir aus (17.9.11) und (17.9.12)

$$E_{Q_0}(n) = \frac{(1-\alpha)\log\dfrac{\beta}{1-\alpha} + \alpha\log\dfrac{1-\beta}{\alpha}}{-\dfrac{1}{2}(Q_1 - Q_0)^2}, \tag{17.9.16}$$

$$E_{Q_1}(n) = \frac{\beta\log\dfrac{\beta}{1-\alpha} + (1-\beta)\log\dfrac{1-\beta}{\alpha}}{\dfrac{1}{2}(Q_1 - Q_0)^2}. \tag{17.9.17}$$

Wir wollen jetzt wieder die Vorteile feststellen, die sich für das hier betrachtete Problem aus der Anwendung eines sequentiellen Quotiententests gegenüber einem klassisch besten Test (siehe 16.3) ergeben. Vor allem interessieren wir uns für die mittlere Anzahl der Beobachtungen.

Die Wahrscheinlichkeiten α und β seien gegeben. Wir wollen die Hypothese $H_0(Q = Q_0)$ gegen die Alternative $H_1(Q = Q_1)$, $Q_1 > Q_0$, testen, wobei Q der unbekannte Mittelwert einer Gesamtheit sei, in der das Merkmal x normal $N(Q; 1)$ verteilt ist. Wenn wir einen klassisch besten Test anwenden, so müssen wir die Hypothese H_0 zugunsten der Hypothese H_1 ablehnen, wenn der aus einer einfachen Stichprobe vom Umfang n ermittelte Wert $\bar{x}$ die Bedingung $\bar{x} \geqq A$ erfüllt, wobei A durch die Formel (16.3.18) gegeben ist. Nach Formel (16.3.17) beträgt der nötige Stichprobenumfang

$$n = \frac{(k_\alpha - k_\beta)^2}{(Q_1 - Q_0)^2},$$

wobei k_α und k_β die Relationen

$$P(y \geqq k_\alpha) = \alpha,$$

$$P(y \leqq k_\beta) = \beta$$

erfüllen, bei denen die Zufallsvariable y die Verteilung $N(0; 1)$ hat.

Wir bezeichnen mit λ_0 und λ_1 die Quotienten aus $E_{Q_0}(n)$ bzw. $E_{Q_1}(n)$ und der durch (16.3.17) gegebenen Zahl n. Dann folgt

$$\lambda_0 = \frac{E_{Q_0}(n)\,(Q_1 - Q_0)^2}{(k_\alpha - k_\beta)^2} = -\frac{2}{(k_\alpha - k_\beta)^2}\left[(1-\alpha)\log\frac{\beta}{1-\alpha} + \alpha\log\frac{1-\beta}{\alpha}\right], \tag{17.9.18}$$

$$\lambda_1 = \frac{E_{Q_1}(n)\,(Q_1 - Q_0)^2}{(k_\alpha - k_\beta)^2} = \frac{2}{(k_\alpha - k_\beta)^2}\left[\beta\log\frac{\beta}{1-\alpha} + (1-\beta)\log\frac{1-\beta}{\alpha}\right]. \tag{17.9.19}$$

Wie wir sehen, hängen λ_0 und λ_1 nur von α und β ab, nicht aber von Q_0 und Q_1, wenn bekannt ist, daß $Q_1 > Q_0$ ist.

Ein ähnliches Ergebnis würden wir erhalten, wenn wir die Hypothese $H_0(Q = Q_0)$ im Vergleich zur Alternative $H_1(Q = Q_1)$ betrachten, wobei $Q_1 < Q_0$ gilt.

Die Tabelle 17.9.1 bringt die Werte des Produkts $100(1 - \lambda_0)$, während in Tabelle 17.9.2 die Werte des Produkts $100(1 - \lambda_1)$ für einige am häufigsten gebrauchte Werte von α und β eingetragen sind.

Tabelle 17.9.1.

$100\,(1 - \lambda_0)$

β \ α	0,01	0,02	0,03	0,04	0,05
0,01	58	54	51	49	47
0,02	60	56	53	50	49
0,03	61	57	54	51	50
0,04	62	58	55	52	50
0,05	63	59	55	53	51

Diese Tabellen geben also für den Fall, daß die entsprechende Hypothese H_0 oder H_1 richtig ist, die durchschnittliche prozentuale Verminderung des Stichprobenumfangs bei Anwendung eines sequentiellen Quotiententests anstelle eines klassisch besten Tests an.

Tabelle 17.9.2.

$100\,(1 - \lambda_1)$

β \ α	0,01	0,02	0,03	0,04	0,05
0,01	58	60	61	62	63
0,02	54	56	57	58	59
0,03	51	53	54	55	55
0,04	49	50	51	52	53
0,05	47	49	50	50	51

Aus diesen Tabellen ist ersichtlich, daß die durchschnittliche prozentuale Verminderung der Beobachtungsanzahl außerordentlich groß ist. Sie beträgt nämlich ungefähr 50%.

17.10. Schlußbemerkungen

Abschließend wollen wir über wichtige Ergebnisse auf dem Gebiet der Sequentialanalyse informieren.

STEIN [2] hat gezeigt, daß beim sequentiellen Quotiententest die Anzahl n der Beobachtungen Momente beliebiger Ordnung besitzt.

WALD und WOLFOWITZ [4] bewiesen den nachstehenden wichtigen Satz, aus dem hervorgeht, daß die in 17.9 bewiesene mittlere Abnahme des Stichprobenumfanges im sequentiellen Quotiententest im Vergleich zu den Testen, die auf Stichproben konstanten Umfangs beruhen, auch auf alle anderen sequentiellen Tests anwendbar ist. Es handelt sich also um eine optimale Eigenschaft, die den sequentiellen Quotiententest charakterisiert. Der Satz lautet:

Satz 17.10.1. *Es sei S_0 ein sequentieller Quotiententest, der zur Überprüfung einer einfachen Hypothese H_0 im Vergleich zur einfachen Alternativhypothese H_1 dient, und es sei S_1 irgendein anderer Test, der demselben Zweck dient. Ferner mögen $\alpha_i(S_j)$ und $E_i^j(n)$ $(i, j = 0, 1)$ die Wahrscheinlichkeit zur Ablehnung der Hypothese H_i bzw. den erwarteten Stichprobenumfang bedeuten, vorausgesetzt, daß die Hypothese H_i gilt und der Test S_j angewandt wird. Darüber hinaus sei $E_i^1(n) < \infty$. Ist dann die Ungleichung*

$$\alpha_i(S_1) \leqq \alpha_i(S_0) \qquad (i = 0, 1)$$

erfüllt, so gilt

$$E_i^0(n) \leqq E_i^1(n).$$

Für das in 17.8 behandelte Problem sind viele Modifikationen des sequentiellen Quotiententests vorgeschlagen worden, die alle zum Ziel hatten, die optimale Eigenschaft beizubehalten, wenn der wahre Wert von Q zwischen Q_0 und Q_1 eingeschlossen ist (siehe ARMITAGE [2], KIEFER und WEISS [1] sowie ANDERSON [2]).

Bei der Behandlung der Theorie des sequentiellen Quotiententests beschränkten wir uns auf den Fall, daß sowohl die Nullhypothese als auch die Alternativhypothese einfach sind. Der sequentielle Quotiententest läßt sich aber auch zur Überprüfung zusammengesetzter Hypothesen anwenden. Informationen zu diesem Thema findet der Leser in dem Buch von WALD [4] und in der Arbeit von GIRSHICK [1].

DVORETZKY, KIEFER und WOLFOWITZ [1] bewiesen, daß die Theorie des sequentiellen Quotiententests sich auf den Fall übertragen läßt, daß es sich bei dem beobachteten Prozeß um einen stochastischen Prozeß mit unabhängigen und homogenen Zuwächsen handelt, und somit insbesondere auf den Fall eines

Poissonschen oder eines Wienerschen Prozesses. In diesem Fall ist die Beobachtungszeit eine Zufallsvariable und spielt eine ähnliche Rolle wie der Stichprobenumfang im sequentiellen Quotiententest. Für den Poissonschen Prozeß, in dem der Gegenstand der Hypothese der Erwartungswert ist, haben KIEFER und WOLFOWITZ [1] spezielle Tabellen berechnet, die mit der Anwendung des sequentiellen Quotiententests zusammenhängen.

Die Sequentialverfahren lassen sich auch auf Probleme der Schätzung anwenden, worauf bereits WALD [4] selbst hingewiesen hat. WOLFOWITZ [2] verallgemeinerte die Ungleichung von RAO-CRAMÉR für den Fall, daß die Anzahl n der Beobachtungen durch ein Sequentialverfahren bestimmt wird, und fand Bedingungen, unter denen man in den Ungleichungen (13.5.1) und (13.5.1') n durch $E(n)$ ersetzen kann. BLACKWELL und GIRSHICK [2] fanden Bedingungen dafür, daß in den auf diese Weise verallgemeinerten Ungleichungen das Zeichen $\geqq$ durch das Gleichheitszeichen ersetzt werden kann. SETH [1] setzte die Untersuchungen von BLACKWELL und GIRSHICK fort. Dem Problem der Gewinnung von Konfidenzintervallen bei Sequentialverfahren ist eine Arbeit von LEIMBACHER [1] gewidmet.

Es sei noch hinzugefügt, daß die Sequentialanalyse nicht nur die Möglichkeiten zur besseren Lösung vieler praktischer Probleme schuf, sondern auch das gesamte Begriffssystem der mathematischen Statistik erheblich bereichert hat.

B. WALD [6] schuf darüber hinaus eine neue allgemeine Theorie, die *Theorie der Entscheidungsfunktionen*, die als Spezialfälle sowohl die klassische Testtheorie für statistische Hypothesen von NEYMAN-PEARSON als auch die Sequentialanalyse als Spezialfälle umfaßt. Die wichtigsten Ideen dieser Theorie lassen sich wie folgt umreißen: Wir bezeichnen mit Ω eine (endliche oder unendliche) Menge der in Betracht kommenden zulässigen Verteilungsfunktionen $F(x)$ einer Zufallsvariablen X und mit ω eine Teilmenge von Ω. Als zulässige Hypothese werden die Hypothesen H_ω angenommen, wobei der Index ω besagt, daß die unbekannte Verteilungsfunktion zu ω gehört. Eine derart formulierte Hypothese kann sowohl eine Konsistenzhypothese als auch eine parametrische Hypothese sein. Die Menge der möglichen Entscheidungen, die in der Theorie von NEYMAN-PEARSON zwei Elemente enthält, d. h. die Entscheidung der Annahme der zu überprüfenden Hypothese und die Entscheidung ihrer Ablehnung zugunsten der Alternativhypothese, die in der Sequentialanalyse erweitert wurde, unterliegt in der neuen Theorie einer weiteren bedeutenden Erweiterung. Dies bezieht sich sowohl auf die möglichen endgültigen Entscheidungen über die Gültigkeit einer der Hypothesen H_ω (solcher Entscheidungen kann es unendlich viele geben) als auch auf Entscheidungen zur Fortsetzung der Untersuchung, wobei es unter diesen letzten Entscheidungen ebenfalls viele geben kann, wenn wir viele Experimente auf einmal durchführen und eine Entscheidung über die Anzahl der fortzusetzenden Experimente treffen müssen. Schließlich werden die Begriffe der Fehler erster und zweiter Art in dieser Theorie verallgemeinert, indem man eine Funktion als

Maß für die Untersuchungskosten und für den wirtschaftlichen Verlust beim Treffen einer falschen Entscheidung einführt.

In der Theorie der Entscheidungsfunktionen wurde ein neuer Apparat zum Testen statistischer Hypothesen entwickelt.

17.11. Aufgaben und Ergänzungen

1. Es sei $\{z_i\}$ $(i = 1, 2, \ldots)$ eine Folge unabhängiger Zufallsvariabler mit gleicher Verteilung. Unter dieser Bedingung gilt die Behauptung des Satzes 17.3.1 dann und nur dann, wenn $P(z_i = 0) < 1$ ist. Man beweise diesen Satz

   a) unter Verwendung des Satzes 17.3.1;

   b) unmittelbar.

2. Man zeige, daß es für die Gültigkeit von (17.6.2) im sequentiellen Quotiententest genügt, anzunehmen, daß $E(n) < \infty$ und daß für jedes Q der Erwartungswert $E_Q(z_i)$ für alle i derselbe ist (Kolmogoroff und Prochorow [1]; siehe auch Blackwell [1]).
   Hinweis. Man wende die Aufgabe 5.14.23 an.

3. Die Wahrscheinlichkeit p $(0 < p < 1)$ für das Auftreten eines fehlerhaften Stücks in einer Serie von Erzeugnissen ist unbekannt. Wir testen die Hypothese $H_0(p = p_0)$ sowie die Alternativhypothese $H_1(p = p_1)$, wobei $p_0 < p_1$ ist. Dabei wenden wir den folgenden Sequentialtest an: Es sei n_0 eine feste ganze Zahl. Wir greifen nacheinander unabhängige Elemente für die Stichprobe heraus. Ist das m-te Stück $(m \leqq n_0)$ fehlerhaft, so lehnen wir H_0 ab; falls jedoch n_0 Stück fehlerfrei sind, wird H_0 akzeptiert.

   a) Man bestimme eine Formel für die Operationscharakteristik dieses Tests.

   b) Wir bezeichnen mit n die Anzahl der Beobachtungen, die zur Entscheidung führt. Man bestimme $E(n)$.

   c) Man setze $p_0 = 0{,}05$, $p_1 = 0{,}1$, $n_0 = 20$ und berechne α, β und $E(n)$.

   d) Man vergleiche die Güte dieses Tests mit der Güte des sequentiellen Quotiententests von 17.8.

4. (Fortsetzung). Man modifiziere den Test der vorhergehenden Aufgabe folgendermaßen: Nach Herausgreifen des m-ten Stücks $(m \leqq n_0 - 1)$ setzen wir die Untersuchung fort, wenn höchstens eines der herausgegriffenen Stücke fehlerhaft ist, und wir lehnen H_0 ab, wenn wenigstens zwei Stück fehlerhaft sind. Nach Herausgreifen von n_0 Stück akzeptieren wir H_0, wenn höchstens ein Stück fehlerhaft ist, und wir lehnen H_0 ab, wenn wenigstens zwei Stücke mit einem Fehler behaftet sind. Man beantworte auch hier alle am Schluß der vorhergehenden Aufgabe gestellten Fragen.

5. Man stelle das Vorgehen bei dem sequentiellen Quotiententest in Form einer Markoffschen Kette dar.

6. Die Zufallsvariable X möge die Verteilung $N(Q; \sigma)$ haben, wobei Q bekannt und σ unbekannt ist. Wir prüfen die Hypothese $H_0(\sigma = \sigma_0)$ im Vergleich zur Alternativhypothese

$H_1(\sigma = \sigma_1)$ mit $\sigma_0 < \sigma_1$. Sodann setzen wir

$$U_m = \sum_{i=1}^{m} (x_i - Q)^2,$$

$$a_m = \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)^{-1} \left(2 \log \frac{\beta}{1-\alpha} + m \log \frac{\sigma_1^2}{\sigma_0^2}\right),$$

$$r_m = \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)^{-1} \left(2 \log \frac{1-\beta}{\alpha} + m \log \frac{\sigma_1^2}{\sigma_0^2}\right).$$

a) Man zeige, daß man beim sequentiellen Quotiententest die Hypothese H_0 annehmen, ablehnen oder ebenfalls die Stichprobenerhebung fortsetzen muß, je nachdem, ob $U_m \leqq a_m$, $U_m \geqq r_m$ oder $a_m < U_m < r_m$ ist (WALD [4]).

b) Man bestimme die Operationscharakteristik dieses Tests und $E(n)$.

7. (Fortsetzung). Man beweise: Ist der Parameter Q ebenfalls unbekannt, so muß in dem oben beschriebenen sequentiellen Quotiententest U_m durch $\sum_{i=1}^{m} (x_i - \bar{x})^2$ ersetzt werden, wobei $\bar{x} = \frac{x_1 + x_2 + \cdots + x_m}{m}$ ist; ferner ist a_m durch a_{m-1} sowie r_m durch r_{m-1} zu ersetzen (GIRSHICK, STEIN).

ANHANG

Der Anhang enthält einen kurzen Abriß über gewisse Elemente der Maßtheorie, die zum Verständnis einiger in diesem Buch benutzter Begriffe benötigt werden. Den Leser, der die allgemeinen Grundlagen der Maßtheorie kennenlernen will, die für die Grundlagen der Wahrscheinlichkeitsrechnung benötigt werden, verweisen wir auf die Bücher von KOLMOGOROFF [7], DOOB [5], HALMOS [2], LOÈVE [4], KAMKE [1], RIESZ und SZ.-NAGY [1] und W. I. SMIRNOW [2].

Im Anhang werden zu den formulierten Sätzen keine Beweise gegeben. Man findet sie in den zitierten Monographien.

Es sei E die Grundmenge (in der Wahrscheinlichkeitsrechnung die Menge der Elementarereignisse). Unter *Mengen* werden wir Teilmengen der Menge E verstehen.

Definition A 1. Es seien A und B Mengen. Die Menge

$$A \dot{-} B = (A - B) \cup (B - A)$$

nennen wir die *symmetrische Differenz* der Mengen A und B.

Definition A 2. Unter einem *Mengenring* (oder *Booleschen Ring*) verstehen wir eine nichtleere Klasse R von Mengen derart, daß aus den Beziehungen

$$A \in R \quad \text{und} \quad B \in R \tag{A 1}$$

die Beziehungen

$$(A \cup B) \in R \quad \text{und} \quad (A - B) \in R \tag{A 2}$$

folgen.

Aus der Definition A 1 geht hervor, daß $(A \dot{-} B) \in R$ gilt, wenn die Beziehungen (A 1) erfüllt sind. Wegen $A \cap B = A \cup B - (A \dot{-} B)$ folgt somit aus (A 1)

$$(A \cap B) \in R. \tag{A 3}$$

Mit Hilfe der vollständigen Induktion finden wir den folgenden Zusammenhang: Ist R ein Mengenring, so folgt für jedes endliche n aus den Beziehungen

$$A_i \in R \qquad (i = 1, 2, \ldots, n) \tag{A 1'}$$

die Beziehung

$$\left(\bigcup_{i=1}^{n} A_i\right) \in R. \tag{A 2'}$$

Beispiel A 1. Wir wählen für E die reelle Zahlengerade und für R die Klasse aller endlichen Vereinigungen beschränkter halboffener Intervalle der Form $[a, b)$. Dann ist R ein Mengenring.

Definition A 3. Einen Mengenring R, der die Grundmenge E enthält, nennen wir eine (*Boolesche*) *Mengenalgebra.*

Ist R eine Mengenalgebra, so folgt aus $A \in R$ die Beziehung $\bar{A} = (E - A) \in R$.

Definition A 4. Unter einem *σ-Ring von Mengen* verstehen wir eine nichtleere Klasse von Mengen derart, daß aus der Beziehung

$$A_i \in R \qquad (i = 1, 2, \ldots) \tag{A 4}$$

die Beziehungen

$$\left(\bigcup_{i=1}^{\infty} A_i\right) \in R, \tag{A 5}$$

$$(A_{i_1} - A_{i_2}) \in R \tag{A 6}$$

folgen.

Definition A 5. Einen σ-Mengenring, der die Grundmenge E enthält, nennen wir eine *σ-Mengenalgebra* (oder einen *Borelschen Mengenkörper*).

Die Definition A 5 ist der folgenden Definition gleichwertig: Unter einer σ-Algebra (bzw. einem Borelschen Mengenkörper) R verstehen wir eine nichtleere Klasse von Mengen derart, daß aus der Bedingung (A 4) die Beziehung (A 5) und aus der Bedingung $A \in R$ die Beziehung $\bar{A} \in R$ folgt.

Wir verweisen den Leser auf die Aufgabe 1.8.6.

Definition A 6. Es sei K eine gewisse Klasse von Mengen. Wir nennen den kleinsten Mengenring (σ-Ring), der die Klasse K enthält, einen *von der Klasse K erzeugten* Ring und bezeichnen ihn mit $g(K)$.

Es läßt sich zeigen, daß zu jeder Klasse K von Mengen ein eindeutig bestimmter Ring (σ-Ring) existiert, der von der Klasse K erzeugt wird.

Beispiel A 2. Wir wählen als E die reelle Zahlengerade und als Klasse K die Klasse aller beschränkten halboffenen Intervalle der Form $[a, b)$. Der von der Klasse K erzeugte Mengenring ist der im Beispiel A 1 betrachtete Ring. Der von der Klasse K erzeugte σ-Ring ist die Klasse aller Borelschen Mengen auf der Zahlengeraden. Dieser σ-Ring ist gleichzeitig eine σ-Algebra, da er die gesamte Gerade als Vereinigung abzählbar vieler Intervalle der Form $[a, b)$ enthält.

Definition A 7. Eine *Mengenfunktion* ist eine Funktion, deren Definitionsbereich eine gewisse Klasse von Mengen ist.

Eine Mengenfunktion μ, die auf einer Klasse K von Mengen definiert ist, nennen wir *endlich additiv*, wenn für $n = 1, 2, \ldots$ und für alle paarweise fremden

Mengen $A_i \in K$ $(i = 1, 2, \ldots, n)$

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i) \tag{A 7}$$

gilt. Wir nennen eine Funktion μ eine *abzählbar additive* (*absolut additive* oder *total additive*) *Funktion*, wenn für jede Folge paarweise disjunkter Mengen $A_i \in K$ $(i = 1, 2, \ldots)$ die Beziehung

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{A 7'}$$

gilt.

Definition A 8. Ein *Maß* μ ist eine reellwertige nichtnegative abzählbar additive Mengenfunktion $\mu(A)$, die auf einem Borelschen Mengenkörper $\mathcal{F}$ definiert ist und der Bedingung $\mu(0) = 0$ genügt. Wir nennen ein Maß μ *endlich*, wenn $\mu(E) < \infty$ ist; ein Maß μ nennen wir ein *Wahrscheinlichkeitsmaß*, wenn $\mu(E) = 1$ ist.

Ein Maß μ nennen wir *σ-endlich*, wenn eine Mengenfolge $\{A_n\}$ im Borelschen Mengenkörper $\mathcal{F}$ derart existiert, daß $E \subset \left(\bigcup_n A_n\right)$ und $\mu(A_n) < \infty$ für $n = 1, 2, \ldots$ gilt.

Das Tripel $(E, \mathcal{F}, \mu)$ nennen wir einen *Maßraum*. Ist $\mu(E) < \infty$, so sagen wir, der Maßraum sei *endlich*. Ist $\mu(E) = 1$, so nennen wir das Tripel $(E, \mathcal{F}, \mu)$ ein *Wahrscheinlichkeitsfeld*.

Der nachstehende wichtige Satz, den man als den *Erweiterungssatz* bezeichnet, zeigt, daß die Voraussetzung, daß das Wahrscheinlichkeitsmaß $P(A)$ für jede Teilmenge A im Borelschen Mengenkörper definiert ist, durch die erheblich schwächere Voraussetzung ersetzt werden kann, daß $P(A)$ für jede Teilmenge A in einer kleineren Mengenalgebra gegeben ist.

Erweiterungssatz. *Es sei $\mu(A)$ ein auf einer gewissen Algebra R von Teilmengen von E definiertes σ-endliches Maß. Dann läßt sich die Funktion $\mu(A)$ in eindeutiger Weise auf alle Mengen aus der Klasse $g(R)$ ausdehnen, ohne daß eine der Eigenschaften der Funktion $\mu(A)$ (nichtnegativ und abzählbar additiv zu sein) verlorengeht.*

Es sei S eine beschränkte Menge auf der reellen Zahlengeraden. Wir wählen $E = (a, b)$, wobei $(a, b) \supset S$ ist. Daraus folgt $\bar{S} = (a, b) - S$. Wir bezeichnen mit U eine endliche oder abzählbare Vereinigungsmenge von Intervallen derart, daß

$$S \subset U \subset (a, b) \tag{A 8}$$

gilt. Wir stellen U als Vereinigung von disjunkten Intervallen dar (siehe Aufgabe 1.8.5) und wählen als Maß $L(U)$ die Summe ihrer Längen.

Definition A 9. Die untere Grenze $L(U)$, erstreckt über alle U, die der Bedingung (A 8) genügen, nennen wir das *äußere Maß der Menge* S. Wir bezeichnen es mit $\bar{L}(S)$.

Der Ausdruck $\underline{L}(S) = b - a - \bar{L}(\bar{S})$ heißt *inneres Maß der Menge* S. Eine Menge S nennen wir *meßbar im Lebesgueschen Sinne* (oder kurz *L-meßbar*), wenn $\underline{L}(S) = \bar{L}(S)$ gilt. Diesen gemeinsamen Wert bezeichnen wir mit $L(S)$.

Definition A 10. Es sei S eine unbeschränkte Menge auf der Zahlengeraden x. Eine Menge S ist *L-meßbar*, wenn für jedes $x > 0$ der Durchschnitt $[-x, x] \cap S$ L-meßbar ist.

Es läßt sich zeigen, daß die Mengenfunktion $L(S)$, die durch die Definitionen A 9 und A 10 gegeben ist, ein auf dem Borelschen Mengenkörper der L-meßbaren Mengen auf der Zahlengeraden definiertes Maß ist, das die folgende Eigenschaft hat: Ist S ein endliches Intervall (a, b), so gilt

$$L(S) = b - a.$$

Jede Borelsche Menge auf der Zahlengeraden ist L-meßbar, doch gilt die umgekehrte Behauptung nicht: Es existieren L-meßbare Mengen, die keine Borelschen Mengen sind. Ist jedoch eine Menge S zwar L-meßbar, aber keine Borelsche Menge, so läßt sie sich in der Form $S = S_1 \cup S_2$ darstellen, wobei S_1 eine Borelsche Menge und S_2 eine L-meßbare Menge mit $L(S_2) = 0$ ist.

Definition A 11. Wir sagen, daß das auf einem Borelschen Mengenkörper $\mathcal{F}$ definierte Maß μ *vollständig* ist, wenn die Bedingungen $A \in \mathcal{F}$, $B \subset A$ und $\mu(A) = 0$ die Beziehung $B \in \mathcal{F}$ nach sich ziehen.

Das Lebesguesche Maß ist vollständig.

Die Verteilungsfunktion einer Zufallsvariablen vom diskreten (stetigen) Typ werden wir als *Verteilungsfunktion vom diskreten (stetigen) Typ* bezeichnen.

Jede Verteilungsfunktion $F(x)$ läßt sich in der Form

$$F(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x) \qquad (a_i \geqq 0;\ a_1 + a_2 + a_3 = 1)$$

darstellen, wobei $F_1(x)$ bzw. $F_2(x)$ Verteilungsfunktionen vom stetigen bzw. diskreten Typ sind, während $F_3(x)$ eine singuläre Verteilungsfunktion ist. Das heißt, $F_3(x)$ ist eine stetige Funktion, und ihre Ableitung $F_3'(x)$ ist fast überall gleich Null (d. h. überall, außer in den Punkten, die zu einer Menge vom Lebesgueschen Maß Null gehören).

Ist $a_2 = a_3 = 0$, so ist die Verteilungsfunktion $F(x)$ vom stetigen Typ. Ist hingegen $a_1 = a_3 = 0$, so ist die Verteilungsfunktion $F(x)$ diskret. Wir geben nun ein Beispiel einer Verteilungsfunktion, für die $a_1 = a_2 = 0$ ist.

Beispiel A 3. Wir teilen das Intervall [0, 1] in die beiden Teile G_0 und T_0 auf folgende Weise:

$$G_0 = \left(\frac{1}{3}, \frac{2}{3}\right) \cap \left[\left(\frac{1}{9}, \frac{2}{9}\right) \cup \left(\frac{7}{9}, \frac{8}{9}\right)\right]$$

$$\cup \left[\left(\frac{1}{27}, \frac{2}{27}\right) \cup \left(\frac{7}{27}, \frac{8}{27}\right) \cup \left(\frac{19}{27}, \frac{20}{27}\right) \cup \left(\frac{25}{27}, \frac{26}{27}\right)\right] \cup \cdots$$

$$\cup \left[\left(\frac{1}{3^n}, \frac{2}{3^n}\right) \cup \left(\frac{7}{3^n}, \frac{8}{3^n}\right) \cup \cdots \cup \left(\frac{3^n - 6}{3^n}, \frac{3^n - 5}{3^n}\right) \cup \left(\frac{3^n - 2}{3^n}, \frac{3^n - 1}{3^n}\right)\right] \cup \cdots,$$

$$T_0 = [0, 1] - G_0.$$

Die Mengen T_0 und G_0 heißen *Cantorsche Mengen.* T_0 ist eine Menge vom Lebesgueschen Maß Null.

Wir bezeichnen mit g_{kn} ($n = 1, 2, \ldots; k = 1, 2, \ldots, 2^{n-1}$) das k-te Intervall mit dem Nenner 3^n. Dann definieren wir $F(x)$ wie folgt:

$$F(x) = \begin{cases} 0 & \text{für} \quad x \leqq 0, \\ \dfrac{2k-1}{2^n} & \text{für} \quad x \in g_{kn}, \\ 1 & \text{für} \quad x \geqq 1. \end{cases}$$

Schließlich erhält man für $x_0 \in T_0$

$$F(x_0) = \sup_{\substack{x < x_0 \\ x \in G_0}} F(x).$$

Das so definierte $F(x)$ ist eine stetige Verteilungsfunktion, deren Ableitung für jedes $x \in G_0$ verschwindet. Es existiert hier demnach keine Funktion $f(x)$, die der Beziehung (2.3.4) für jedes reelle x genügt.

Es ist wichtig, festzustellen, daß bei unabhängigen Zufallsvariablen X_1 und X_2, deren Verteilungsfunktionen $F_1(x)$ und $F_2(x)$ singulär sind, die Verteilungsfunktion $F(x)$ der Summe $X_1 + X_2$ absolut stetig sein kann (siehe Ranga-Rao und Varadarajan [1]).

Es sei $(E, \mathcal{F}, \mu)$ ein endlicher Maßraum. Ferner sei $f(e)$, $e \in E$, eine reelle, bezüglich $\mathcal{F}$ meßbare (vgl. Definition 2.1.2 sowie die zugehörige Fußnote) und beschränkte Funktion: $-\infty < a \leqq f(e) \leqq b < \infty$. Wir bezeichnen mit $\mu\,(c' \leqq f(e) < c'', e \in A)$, wobei $A \in \mathcal{F}$ ist, das Maß des Durchschnitts von A mit dem Urbild von (c', c''), aus dem (c', c'') durch die Funktion f entstanden ist.

Definition A 12. Existiert der Grenzwert des Ausdrucks

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \sum_{i=1}^{n} c_i \mu(c_{i-1} \leqq f < c_i, e \in A), \tag{A 9}$$

wobei $a = c_0 < c_1 < \cdots < c_n = b$ und $\max_{1 \leqq i \leqq n} (c_i - c_{i-1}) \leqq \varepsilon$ ist, so nennen wir diesen Grenzwert das *abstrakte Lebesguesche Integral von f über der Menge A bezüglich des Maßes μ*. Wir bezeichnen es wie folgt:

$$\int_A f \, d\mu .$$

Ist die Funktion $f(e)$ beschränkt, so existiert das Integral für jedes $A \in \mathcal{F}$ und insbesondere für $A = E$. Falls die Funktion $f(e)$ unbeschränkt ist, verfahren wir wie folgt: Wir setzen für jedes a und b mit $a < b$

$$f_{ab}(e) = \begin{cases} a & \text{für} \quad f(e) \leqq a, \\ f(e) & \text{für} \quad a < f(e) < b, \\ b & \text{für} \quad f(e) \geqq b. \end{cases}$$

Definition A 12′. Existiert der Grenzwert

$$\lim_{\substack{a \to -\infty \\ b \to \infty}} \int_A f_{ab}(e) \, d\mu , \tag{A 9′}$$

so nennen wir ihn das *abstrakte Lebesguesche Integral der Funktion f über der Menge A bezüglich des Maßes μ.*

Wir wollen nun einige wichtige Eigenschaften des Lebesgueschen Integrals anführen.

I. Ist die Funktion f über der Menge A integrierbar und ist $A = \sum_n A_n$ ($A_n \in \mathcal{F}$; $n = 1, 2, \ldots$), wobei die Mengen A_n paarweise disjunkt sind, so gilt

$$\int_A f \, d\mu = \sum_n \int_{A_n} f \, d\mu .$$

II. Ist die Funktion f integrierbar, so ist auch $|f|$ eine integrierbare Funktion.

III. Sind f_1 und f_2 integrierbare Funktionen und c_1 und c_2 reelle Konstanten, so ist auch die Funktion $c_1 f_1 + c_2 f_2$ integrierbar, und es gilt

$$\int_A (c_1 f_1 + c_2 f_2) \, d\mu = c_1 \int_A f_1 d\mu + c_2 \int_A f_2 d\mu .$$

Wir wählen jetzt als μ das Wahrscheinlichkeitsmaß P; f sei die charakteristische Funktion der Menge $A \in \mathcal{F}$, d. h., es sei

$$f(e) = \begin{cases} 1 & \text{für} \quad e \in A, \\ 0 & \text{für} \quad e \notin A. \end{cases}$$

Dann gilt

$$\int_A f \, dP = P(A).$$

Nun sei X eine Zufallsvariable mit der Verteilungsfunktion $F(x)$, und S sei eine Borelsche Menge über der reellen Zahlengeraden. Dann ist

$$P^{(x)}(S) = P^{(x)}(X \in S) = \int_S dF(x) = \int_A dP = P(A),$$

wobei A das Urbild von S ist.

Eine völlig analoge Formel gilt für Zufallsvektoren in n-dimensionalen euklidischen Räumen.

Existiert der Erwartungswert $E(X)$, so gilt

$$E(X) = \int_E x\,dP = \int_{-\infty}^{\infty} x\,dF(x) = \int_{-\infty}^{\infty} x\,dP^{(x)}.$$

Ist $X = (X_1, X_2, \ldots, X_n)$ ein Zufallsvektor und $g(X_1, X_2, \ldots, X_n)$ eine Zufallsvariable, deren Erwartungswert existiert, so ist

$$E[g(X_1, \ldots, X_n)] = \int_E g\,dP = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g\,dP^{(x_1, \ldots, x_n)}.$$

Bevor wir den Satz von Fubini formulieren, führen wir einige Begriffe ein.

Definition A 13. Es seien $(E_i, \mathcal{F}_i, \mu_i)$ $(i = 1, 2)$ zwei Maßräume. Das Tripel $(E, \mathcal{F}, \mu)$ nennen wir das *kartesische Produkt der Maßräume* $(E_1, \mathcal{F}_1, \mu_1)$ *und* $(E_2, \mathcal{F}_2, \mu_2)$, wenn die nachstehenden Bedingungen erfüllt sind.

I. $E = E_1 \times E_2$ ist die Menge aller Paare $e = (e_1, e_2)$ mit $e_i \in E_i$.

II. $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$ ist der von den Rechtecken $A = A_1 \times A_2$ erzeugte Borelsche Mengenkörper, wobei $A_i \in \mathcal{F}_i$ ist (mit anderen Worten, $e \in A$ bedeutet, daß $e = (e_1, e_2)$ gilt, sofern $e_i \in A_i$ ist).

III. Für jedes $A = A_1 \times A_2$ mit $A_i \in \mathcal{F}_i$ haben wir $\mu(A) = \mu_1(A_1)\,\mu_2(A_2)$. Das Maß μ nennen wir das *kartesische Produkt der Maße* μ_1 *und* μ_2 und bezeichnen es mit $\mu_1 \times \mu_2$.

Beispiel A 4. Es seien X und Y zwei unabhängige Zufallsvariable. Wir wählen die x-Achse, die y-Achse und die x,y-Ebene als die Mengen E_1, E_2 und E sowie die Klassen der Borelschen Mengen $\mathcal{F}_x$, $\mathcal{F}_y$ und $\mathcal{F}_{(x,y)}$ über den Mengen E_1, E_2 und E als $\mathcal{F}_1$, $\mathcal{F}_2$ und $\mathcal{F}$ Für die Maße μ_1, μ_2 und μ schließlich wählen wir die Wahrscheinlichkeitsmaße $P^{(x)}$, $P^{(y)}$ und $P^{(x,y)}$ der Zufallsvariablen X, Y und (X, Y). Dann ist $(E, \mathcal{F}_{(x,y)}, P^{(x,y)})$ das kartesische Produkt der Maßräume $(E_1, \mathcal{F}_x, P^{(x)})$ und $(E_2, \mathcal{F}_y, P^{(y)})$.

Definition A 13'. Es sei $A \in \mathcal{F}$ $(A = A_1 \times A_2)$. Als *Schnitt der Menge* A *in einem gegebenen Element* $e_1(e_2)$ bezeichnen wir die Menge aller Elemente $e_2(e_1)$ derart, daß $(e_1, e_2) \in A$ ist. Für diesen Schnitt schreiben wir $A_{e_1}(A_{e_2})$.

Es sei f eine auf der Menge E erklärte, bezüglich $\mathcal{F}$ meßbare reelle Funktion. Als Schnitt $f_{e_1}(f_{e_2})$ der Funktion f im gegebenen Element $e_1(e_2)$ bezeichnen wir die Funktion, die auf der Menge $E_2(E_1)$ durch die Beziehung $f_{e_1}(e_2) = f(e_1, e_2)$ (oder $f_{e_2}(e_1) = f(e_1, e_2)$) erklärt ist.

Satz von Fubini. *Es seien* $(E_i, \mathcal{F}_i, \mu_i)$ $(i = 1, 2)$ *zwei endliche Maßräume. Ist die auf der Menge* $E_1 \times E_2$ *definierte Funktion* $f(e_1, e_2)$ *bezüglich* $\mathcal{F}_1 \times \mathcal{F}_2$ *meßbar und nichtnegativ oder integrierbar auf der Menge* $E = E_1 \times E_2$ *nach dem Maß* $\mu = \mu_1 \times \mu_2$, *so ist*

$$\int_E f\, d\mu = \int_{E_1} d\mu_1 \int_{E_2} f_{e_1}(e_2)\, d\mu_2 = \int_{E_2} d\mu_2 \int_{E_1} f_{e_2}(e_1)\, d\mu_1. \tag{A 10}$$

Im Fall der Integrabilität sind fast alle Schnitte (eventuell mit Ausnahme einer Menge vom Maß Null) integrierbar.

Die Integrale auf der rechten Seite von (A 10) sind der Reihe nach von rechts nach links zu berechnen.

Ist f charakteristische Funktion der Menge $A \in \mathcal{F}$, so nimmt die Beziehung (A 10) die folgende Form an:

$$\mu(A) = \int_{E_1} \mu_2(A_{e_1})\, d\mu_1 = \int_{E_2} \mu_1(A_{e_2})\, d\mu_2. \tag{A 11}$$

Der Satz von Fubini gestattet uns, einen einfachen Beweis für die Formeln (6.5.6) bis (6.5.9) zu geben. Wir wollen etwa die Formel (6.5.6) beweisen.

Es seien X und Y unabhängige Zufallsvariable; $Z = f(X, Y)$ sei ebenfalls eine Zufallsvariable; ihr Wahrscheinlichkeitsmaß sei $P^{(z)}$. Ferner sei S eine Borelsche Menge auf der z-Achse und A eine Menge in der x, y-Ebene mit der folgenden Eigenschaft: Für $(x, y) \in A$ ist $f(x, y) \in S$. Aus der Beziehung (A 11) folgt dann (vgl. Beispiel A 4) wegen $P^{(x,y)}(A) = P^{(z)}(S)$

$$P^{(z)}(S) = \int_{-\infty}^{\infty} P^{(y)}(A_x)\, dP^{(x)} = \int_{-\infty}^{\infty} P^{(x)}(A_y)\, dP^{(y)}; \tag{A 12}$$

dabei ist A_x (A_y) die Menge aller der Punkte auf der y-Achse (x-Achse), bei denen für ein gegebenes $x\,(y)$ die Beziehung $f(x, y) \in S$ gilt.

Wir bezeichnen mit $F_1(x)$, $F_2(y)$ bzw. $F(z)$ die Verteilungsfunktionen der Zufallsvariablen X, Y bzw. Z. Setzen wir $S = (-\infty, z)$ und $Z = X + Y$, so folgt

$$P^{(z)}(S) = F(z), \quad P^{(y)}(A_x) = F_2(z - x), \quad P^{(x)}(A_y) = F_1(z - y). \tag{A 13}$$

Aus (A 12) und (A 13) erhalten wir sofort die Formel (6.5.6). Die übrigen drei Formeln lassen sich in der gleichen Weise gewinnen.

Mit Hilfe des Satzes von Fubini läßt sich auch die nachstehende Umkehrung zu Satz 3.6.2 beweisen.

Es seien X und Y unabhängige Zufallsvariable, die die Beziehungen

$$P(X=0) \neq 1, \qquad P(Y=0) \neq 1 \tag{A 14}$$

erfüllen, und es existiere der Erwartungswert $E(XY)$. Dann existieren auch die Erwartungswerte $E(X)$ und $E(Y)$, und es gilt

$$E(XY) = E(X)\,E(Y). \tag{A 15}$$

Zum Beweis nehmen wir als Funktion f aus der Beziehung (A 10) das Produkt XY. Der Schnitt der Funktion f in einem gegebenen Punkt x ist xy bei konstantem x; beachtet man die Beziehung (A 14), so ergibt sich aus (A 10)

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,dP^{(x,y)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\,dP^{(y)}\,dP^{(x)} = E(X)\,E(Y).$$

Die Existenz und Eindeutigkeit der bedingten Erwartungswerte und der bedingten Wahrscheinlichkeit, wie sie in 3.6.C besprochen wurden, folgen leicht aus dem Satz von Radon-Nikodym.

Es sei $(E, \mathcal{F}, \mu)$ ein Maßraum und $V(A)$ eine abzählbar additive Mengenfunktion, die für $A \in \mathcal{F}$ definiert ist. Ferner möge $V(A)$ höchstens einen der Werte $-\infty$ und ∞ annehmen.

Wir sagen, daß die Mengenfunktion V *absolut stetig bezüglich des Maßes* μ ist (wir schreiben dafür: $V \ll \mu$), wenn für jedes $A \in \mathcal{F}$ die Gleichung $\mu(A) = 0$ die Beziehung $V(A) = 0$ nach sich zieht.

Ist die Funktion V endlich, nichtnegativ und ist $V \ll \mu$, so existiert zu jedem $\varepsilon > 0$ ein $\delta > 0$ derart, daß die Beziehung $\mu(A) > \delta$ die Beziehung $V(A) < \varepsilon$ nach sich zieht.

Satz von Radon-Nikodym. *Es sei $(E, \mathcal{F}, P)$ ein Wahrscheinlichkeitsfeld und $Q(A)$ $(A \in \mathcal{F})$ eine abzählbar additive endliche Mengenfunktion. Ist $Q \ll P$, so existiert eine Zufallsvariable $f(e)$ derart, daß für jedes $A \in \mathcal{F}$ die Beziehung*

$$Q(A) = \int_A f(e)\,dP \tag{A 16}$$

erfüllt ist.

Die Zufallsvariable $f(e)$ ist in dem Sinne eindeutig bestimmt, daß dann, wenn für eine andere Zufallsvariable $g(e)$ für jedes $A \in \mathcal{F}$ die Beziehung

$$Q(A) = \int_A g(e)\,dP$$

gilt, die Beziehung $P[f(e) = g(e)] = 1$ folgt.

Es sei (X, Y) ein Zufallsvektor, und es möge der Erwartungswert $E(Y)$ existieren.

Definition A 14. Die Zufallsvariable $E(Y|x)$, die für jede Borelsche Menge S auf der x-Achse, für die $P^{(x)}(X \in S) > 0$ gilt, die Beziehung

$$E(Y|X \in S) = E[E(Y|x)|X \in S] \tag{A 17}$$

erfüllt, nennen wir den *bedingten Erwartungswert* der Zufallsvariablen Y bei gegebenem x.

Wir beweisen die Existenz und Eindeutigkeit (in dem im Satz von Radon-Nikodym formulierten Sinne) des bedingten Erwartungswertes $E(Y|x)$.

Es ist hier zweckmäßig, mit $\{X \in S\}$ das Urbild der Borelschen Menge S zu bezeichnen, das durch die Zufallsvariable X gegeben ist. Multiplizieren wir beide Seiten von (A 17) mit $P(X \in S)$, so erhalten wir

$$\int_{\{X \in S\}} y\,dP = \int_{\{X \in S\}} E(Y|x)\,dP = \int_S E(Y|x)\,dP^{(x)}. \tag{A 18}$$

Nun sei

$$Q(S) = \int_{\{X \in S\}} y\,dP.$$

Die Funktion $Q(S)$ ist offenbar abzählbar additiv. Darüber hinaus ist $Q(S)$ absolut stetig nach dem Maß $P^{(x)}$, denn die Gleichung $P^{(x)}(S) = 0$ ist der Gleichung $P(X \in S) = 0$ äquivalent, und aus dieser folgt $Q(S) = 0$.

Wie wir sehen, sind die Voraussetzungen des Satzes von Radon-Nikodym erfüllt, wenn wir für die Menge E die x-Achse, für $\mathcal{F}$ die Klasse der Borelschen Mengen auf dieser Achse und für P das Maß $P^{(x)}$ wählen. Demzufolge ist die Behauptung des Satzes von Radon-Nikodym für $E(Y|x)$ zutreffend.

TAFELN

Tafel I. *Die Poissonsche Verteilung*
Die Werte der Wahrscheinlichkeitsfunktion

$$P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda}$$

einer Zufallsvariablen X mit Poissonscher Verteilung

r	λ							
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812	0,496585	0,449329
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287	0,347610	0,359463
2	0,004524	0,016375	0,033337	0,053626	0,075816	0,098786	0,121663	0,143785
3	0,000151	0,001092	0,003334	0,007150	0,012636	0,019757	0,028388	0,038343
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964	0,004968	0,007669
5	—	0,000002	0,000015	0,000057	0,000158	0,000356	0,000696	0,001227
6	—	—	0,000001	0,000004	0,000013	0,000036	0,000081	0,000164
7	—	—	—	—	0,000001	0,000003	0,000008	0,000019
8	—	—	—	—	—	—	0,000001	0,000002

r	λ							
	0,9	1,0	1,5	2,0	2,5	3,0	3,5	4,0
0	0,406570	0,367879	0,223130	0,135335	0,082085	0,049787	0,030197	0,018316
1	0,365913	0,367879	0,334695	0,270671	0,205212	0,149361	0,105691	0,073263
2	0,164661	0,183940	0,251021	0,270671	0,256516	0,224042	0,184959	0,146525
3	0,049398	0,061313	0,125510	0,180447	0,213763	0,224042	0,215785	0,195367
4	0,011115	0,015328	0,047067	0,090224	0,133602	0,168031	0,188812	0,195367
5	0,002001	0,003066	0,014120	0,036089	0,066801	0,100819	0,132169	0,156293
6	0,000300	0,000511	0,003530	0,012030	0,027834	0,050409	0,077098	0,104196
7	0,000039	0,000073	0,000756	0,003437	0,009941	0,021604	0,038549	0,059540
8	0,000004	0,000009	0,000142	0,000859	0,003106	0,008102	0,016865	0,029770
9	—	0,000001	0,000024	0,000191	0,000863	0,002701	0,006559	0,013231
10	—	—	0,000004	0,000038	0,000216	0,000810	0,002296	0,005292
11	—	—	—	0,000007	0,000049	0,000221	0,000730	0,001925
12	—	—	—	0,000001	0,000010	0,000055	0,000213	0,000642
13	—	—	—	—	0,000002	0,000013	0,000057	0,000197
14	—	—	—	—	—	0,000003	0,000014	0,000056
15	—	—	—	—	—	0,000001	0,000003	0,000015
16	—	—	—	—	—	—	0,000001	0,000004
17	—	—	—	—	—	—	—	0,000001

Fortsetzung von Tafel I

r	λ 4,5	5,0	6,0	7,0	8,0	9,0	10,0
0	0,011109	0,006738	0,002479	0,000912	0,000335	0,000123	0,000045
1	0,049990	0,033690	0,014873	0,006383	0,002684	0,001111	0,000454
2	0,112479	0,083224	0,044618	0,022341	0,010735	0,004998	0,002270
3	0,168718	0,140374	0,089235	0,052129	0,028626	0,014994	0,007567
4	0,189808	0,175467	0,133853	0,091226	0,057252	0,033737	0,018917
5	0,170827	0,175467	0,160623	0,127717	0,091604	0,060727	0,037833
6	0,128120	0,146223	0,160623	0,149003	0,122138	0,091090	0,063055
7	0,082363	0,104445	0,137677	0,149003	0,139587	0,117116	0,090079
8	0,046329	0,065278	0,103258	0,130377	0,139587	0,131756	0,112599
9	0,023165	0,036266	0,068838	0,101405	0,124077	0,131756	0,125110
10	0,010424	0,018133	0,041303	0,070983	0,099262	0,118580	0,125110
11	0,004264	0,008242	0,022529	0,045171	0,072190	0,097020	0,113736
12	0,001599	0,003434	0,011264	0,026350	0,048127	0,072765	0,094780
13	0,000554	0,001321	0,005199	0,014188	0,029616	0,050376	0,072908
14	0,000178	0,000472	0,002228	0,007094	0,016924	0,032384	0,052077
15	0,000053	0,000157	0,000891	0,003311	0,009026	0,019431	0,034718
16	0,000015	0,000049	0,000334	0,001448	0,004513	0,010930	0,021699
17	0,000004	0,000014	0,000118	0,000596	0,002124	0,005786	0,012764
18	0,000001	0,000004	0,000039	0,000232	0,000944	0,002893	0,007091
19	—	0,000001	0,000012	0,000085	0,000397	0,001370	0,003732
20	—	—	0,000004	0,000030	0,000159	0,000617	0,001866
21	—	—	0,000001	0,000010	0,000061	0,000264	0,000889
22	—	—	—	0,000003	0,000022	0,000108	0,000404
23	—	—	—	0,000001	0,000008	0,000042	0,000176
24	—	—	—	—	0,000003	0,000016	0,000073
25	—	—	—	—	0,000001	0,000006	0,000029
26	—	—	—	—	—	0,000002	0,000011
27	—	—	—	—	—	0,000001	0,000004
28	—	—	—	—	—	—	0,000001
29	—	—	—	—	—	—	0,000001

Tafel II. *Die Normalverteilung*

Die Werte der Dichte

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

einer Zufallsvariablen X mit der Normalverteilung $N(0;1)$

x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$
0,00	0,3989	0,50	0,3521	1,00	0,2420	1,50	0,1295	2,00	0,0540	2,50	0,0175
0,05	0,3984	0,55	0,3429	1,05	0,2299	1,55	0,1200	2,05	0,0488	2,55	0,0154
0,10	0,3970	0,60	0,3332	1,10	0,2179	1,60	0,1109	2,10	0,0440	2,60	0,0136
0,15	0,3945	0,65	0,3230	1,15	0,2059	1,65	0,1023	2,15	0,0396	2,65	0,0119
0,20	0,3910	0,70	0,3123	1,20	0,1942	1,70	0,0940	2,20	0,0355	2,70	0,0104
0,25	0,3867	0,75	0,3011	1,25	0,1826	1,75	0,0863	2,25	0,0317	2,75	0,0091
0,30	0,3814	0,80	0,2897	1,30	0,1714	1,80	0,0790	2,30	0,0283	2,80	0,0079
0,35	0,3752	0,85	0,2780	1,35	0,1604	1,85	0,0721	2,35	0,0252	2,85	0,0069
0,40	0,3683	0,90	0,2661	1,40	9,1497	1,90	0,0656	2,40	0,0224	2,90	0,0060
0,45	0,3605	0,95	0,2541	1,45	0,1394	1,95	0,0596	2,45	0,0198	2,95	0,0051
										3,00	0,0044

Tafel III. *Die Normalverteilung*

Die Werte der Verteilungsfunktion

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

einer Zufallsvariablen X mit der Normalverteilung $N(0;1)$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,500000	0,50	0,691463	1,00	0,841345	1,50	0,933193	2,00	0,977250	2,50	0,993790
0,05	0,519939	0,55	0,708840	1,05	0,853141	1,55	0,939429	2,05	0,979818	2,55	0,994614
0,10	0,539828	0,60	0,725747	1,10	0,864334	1,60	0,945201	2,10	0,982136	2,60	0,995339
0,15	0,559618	0,65	0,742154	1,15	0,874928	1,65	0,950528	2,15	0,984222	2,65	0,995975
0,20	0,579260	0,70	0,758036	1,20	0,884930	1,70	0,955434	2,20	0,986097	2,70	0,996533
0,25	0,589706	0,75	0,773373	1,25	0,894350	1,75	0,959941	2,25	0,987776	2,75	0,997020
0,30	0,617911	0,80	0,788145	1,30	0,903200	1,80	0,964070	2,30	0,989276	2,80	0,997445
0,35	0,636831	0,85	0,802338	1,35	0,911492	1,85	0,967843	2,35	0,990613	2,85	0,997814
0,40	0,655422	0,90	0,815940	1,40	0,919243	1,90	0,971283	2,40	0,991802	2,90	0,998134
0,45	0,673645	0,95	0,828944	1,45	0,926471	1,95	0,974412	2,45	0,992857	2,95	0,998411
										3,00	0,998650

Tafel IV. *Die χ^2-Verteilung*

Die Tafel gibt die Werte von χ^2_α für einige Werte α an. Dabei ist χ^2_α so bestimmt, daß die Wahrscheinlichkeit dafür, daß die Zufallsvariable χ^2 mit n Freiheitsgraden nicht kleiner als χ^2_α ist, gleich α ist:

$$P(\chi^2 \geqq \chi^2_\alpha) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int\limits_{\chi^2_\alpha}^{\infty} e^{-\frac{u}{2}} u^{\frac{n}{2}-1}\, du = \alpha$$

n	α								
	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,064	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345
4	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,622	18,475
8	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

Tafel V. *Die Studentsche t-Verteilung*

Die Tafel enthält die Werte von t_α für einige Werte α. Dabei ist t_α derart gewählt, daß die Wahrscheinlichkeit dafür, daß die Studentsche Zufallsvariable t mit n Freiheitsgraden absolut genommen nicht kleiner als t_α ist, gleich α ist:

$$P(|t| \geqq t_\alpha) = \frac{2}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{t_\alpha}^{\infty} \frac{1}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}\, dt = \alpha$$

n	α							
	0,80	0,60	0,40	0,20	0,10	0,05	0,02	0,01
1	0,325	0,727	1,376	3,078	6,314	12,706	31,821	63,657
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106
12	0,259	0,539	0,873	1 356	1,782	2,179	2,681	3,055
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750
40	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660
120	0,254	0,526	0,845	1,289	1,658	1,980	2,358	2,617
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576

Tafel VI. *Die Fishersche Z-Verteilung*

Die Tafel enthält die Werte von z_0, für die die Wahrscheinlichkeit dafür, daß die Fishersche Zufallsvariable Z mit (r_1, r_2) Freiheitsgraden nicht kleiner als z_0 ist, gleich 0,05 ist:

$$P(Z \geqq z_0) = \int_{z_0}^{\infty} f(z)\, dz = 0{,}05,$$

hierbei ist $f(z)$ durch die Formel (9.7.5) bestimmt.

r_2	r_1									
	1	2	3	4	5	6	8	12	24	∞
1	2,5421	2,6479	2,6870	2,7071	2,7194	2,7276	2,7380	2,7484	2,7588	2,7693
2	1,4592	1,4722	1,4765	1,4787	1,4800	1,4808	1,4819	1,4830	1,4840	1,4851
3	1,1577	1,1284	1,1137	1,1051	1,0994	1,0953	1,0899	1,0842	1,0781	1,0716
4	1,0212	0,9690	0,9429	0,9272	0,9168	0,9093	0,8993	0,8885	0,8767	0,8639
5	0,9441	0,8777	0,8441	0,8236	0,8097	0,7997	0,7862	0,7714	0,7550	0,7368
6	0,8948	0,8188	0,7798	0,7558	0,7394	0,7274	0,7112	0,6931	0,6729	0,6499
7	0,8606	0,7777	0,7347	0,7080	0,6896	0,6761	0,6576	0,6369	0,6134	0,5862
8	0,8355	0,7475	0,7014	0,6725	0,6525	0,6378	0,6175	0,5945	0,5682	0,5371
9	0,8163	0,7242	0,6757	0,6450	0,6238	0,6080	0,5862	0,5613	0,5324	0,4979
10	0,8012	0,7058	0,6553	0,6232	0,6009	0,5843	0,5611	0,5346	0,5035	0,4657
11	0,7889	0,6909	0,6387	0,6055	0,5822	0,5648	0,5406	0,5126	0,4795	0,4387
12	0,7788	0,6786	0,6250	0,5907	0,5666	0,5487	0,5234	0,4941	0,4592	0,4156
13	0,7703	0,6682	0,6134	0,5783	0,5535	0,5350	0,5089	0,4785	0,4419	0,3957
14	0,7630	0,6594	0,6036	0,5677	0,5423	0,5233	0,4964	0,4649	0,4269	0,3782
15	0,7568	0,6518	0,5950	0,5585	0,5326	0,5131	0,4855	0,4532	0,4138	0,3628
16	0,7514	0,6451	0,5876	0,5505	0,5241	0,5042	0,4760	0,4428	0,4022	0,3490
17	0,7466	0,6393	0,5811	0,5434	0,5166	0,4964	0,4676	0,4337	0,3919	0,3366
18	0,7424	0,6341	0,5753	0,5371	0,5099	0,4894	0,4602	0,4255	0,3827	0,3253
19	0,7386	0,6295	0,5701	0,5315	0,5040	0,4832	0,4535	0,4182	0,3743	0,3151
20	0,7352	0,6254	0,5654	0,5265	0,4986	0,4776	0,4474	0,4116	0,3668	0,3057
21	0,7322	0,6216	0,5612	0,5219	0,4938	0,4725	0,4420	0,4055	0,3599	0,2971
22	0,7294	0,6182	0,5574	0,5178	0,4894	0,4679	0,4370	0,4001	0,3536	0,2892
23	0,7269	0,6151	0,5540	0,5140	0,4854	0,4636	0,4325	0,3950	0,3478	0,2818
24	0,7246	0,6123	0,5508	0,5106	0,4817	0,4598	0,4283	0,3904	0,3425	0,2749
25	0,7225	0,6097	0,5478	0,5074	0,4783	0,4562	0,4244	0,3862	0,3376	0,2685
26	0,7205	0,6073	0,5451	0,5045	0,4752	0,4529	0,4209	0,3823	0,3330	0,2625
27	0,7187	0,6051	0,5427	0,5017	0,4723	0,4499	0,4176	0,3786	0,3287	0,2569
28	0,7171	0,6030	0,5403	0,4992	0,4696	0,4471	0,4146	0,3752	0,3248	0,2516
29	0,7155	0,6011	0,5382	0,4969	0,4671	0,4444	0,4117	0,3720	0,3211	0,2466
30	0,7141	0,5994	0,5362	0,4947	0,4648	0,4420	0,4090	0,3691	0,3176	0,2419
40	0,7037	0,5866	0,5217	0,4789	0,4479	0,4242	0,3897	0,3475	0,2920	0,2057
60	0,6933	0,5738	0,5073	0,4632	0,4311	0,4064	0,3702	0,3255	0,2654	0,1644
120	0,6830	0,5611	0,4930	0,4475	0,4143	0,3885	0,3506	0,3032	0,2376	0,1131
∞	0,6729	0,5486	0,4787	0,4319	0,3974	0,3706	0,3309	0,2804	0,2085	0,0000

Tafel VII. *Die Fishersche Z-Verteilung*

Die Tafel enthält die Werte von z_0, für die die Wahrscheinlichkeit dafür, daß die Fishersche Zufallsvariable Z mit (r_1, r_2) Freiheitsgraden nicht kleiner als z_0 ist, gleich 0,01 ist.

$$P(Z \geqq z_0) = \int_{z_0}^{\infty} f(z)\, dz = 0{,}01,$$

hierbei ist $f(z)$ durch die Formel (9.7.5) bestimmt.

r_2	r_1									
	1	2	3	4	5	6	8	12	24	∞
1	4,1535	4,2585	4,2974	4,3175	4,3297	4,3379	4,3482	4,3585	4,3689	4,3794
2	2,2950	2,2976	2,2984	2,2988	2,2991	2,2992	2,2994	2,2997	2,2999	2,3001
3	1,7649	1,7140	1,6915	1,6786	1,6703	1,6645	1,6569	1,6489	1,6404	1,6314
4	1,5270	1,4452	1,4075	1,3856	1,3711	1,3609	1,3473	1,3327	1,3170	1,3000
5	1,3943	1,2929	1,2449	1,2164	1,1974	1,1838	1,1656	1,1457	1,1239	1,0997
6	1,3103	1,1955	1,1401	1,1068	1,0843	1,0680	1,0460	1,0218	0,9948	0,9643
7	1,2526	1,1281	1,0682	1,0300	1,0048	0,9864	0,9614	0,9335	0,9020	0,8658
8	1,2106	1,0787	1,0135	0,9734	0,9459	0,9259	0,8983	0,8673	0,8319	0,7904
9	1,1786	1,0411	0,9724	0,9299	0,9006	0,8791	0,8494	0,8157	0,7769	0,7305
10	1,1535	1,0114	0,9399	0,8954	0,8646	0,8419	0,8104	0,7744	0,7324	0,6816
11	1,1333	0,9874	0,9136	0,8674	0,8354	0,8116	0,7785	0,7405	0,6958	0,6408
12	1,1166	0,9677	0,8919	0,8443	0,8111	0,7864	0,7520	0,7122	0,6649	0,6061
13	1,1027	0,9511	0,8737	0,8248	0,7907	0,7652	0,7295	0,6882	0,6386	0,5761
14	1,0909	0,9370	0,8581	0,8082	0,7732	0,7471	0,7103	0,6675	0,6159	0,5500
15	1,0807	0,9249	0,8448	0,7939	0,7582	0,7314	0,6937	0,6496	0,5961	0,5269
16	1,0719	0,9144	0,8331	0,7814	0,7450	0,7177	0,6791	0,6339	0,5786	0,5064
17	1,0641	0,9051	0,8229	0,7705	0,7335	0,7057	0,6663	0,6199	0,5630	0,4879
18	1,0572	0,8970	0,8138	0,7607	0,7232	0,6950	0,6549	0,6075	0,5491	0,4712
19	1,0511	0,8897	0,8057	0,7521	0,7140	0,6854	0,6447	0,5964	0,5366	0,4560
20	1,0457	0,8831	0,7985	0,7443	0,7058	0,6768	0,6355	0,5864	0,5253	0,4421
21	1,0408	0,8772	0,7920	0,7372	0,6984	0,6690	0,6272	0,5773	0,5150	0,4294
22	1,0363	0,8719	0,7860	0,7309	0,6916	0,6620	0,6196	0,5691	0,5056	0,4176
23	1,0322	0,8670	0,7806	0,7251	0,6855	0,6555	0,6127	0,5615	0,4969	0,4068
24	1,0285	0,8626	0,7757	0,7197	0,6799	0,6496	0,6064	0,5545	0,4890	0,3967
25	1,0251	0,8585	0,7712	0,7148	0,6747	0,6442	0,6006	0,5481	0,4816	0,3872
26	1,0220	0,8548	0,7670	0,7103	0,6699	0,6392	0,5952	0,5422	0,4748	0,3784
27	1,0191	0,8513	0,7631	0,7062	0,6655	0,6346	0,5902	0,5367	0,4685	0,3701
28	1,0164	0,8481	0,7595	0,7023	0,6614	0,6303	0,5856	0,5316	0,4626	0,3624
29	1,0139	0,8451	0,7562	0,6987	0,6576	0,6263	0,5813	0,5269	0,4570	0,3550
30	1,0116	0,8423	0,7531	0,6954	0,6540	0,6226	0,5773	0,5224	0,4519	0,3481
40	0,9949	0,8223	0,7307	0,6712	0,6283	0,5956	0,5481	0,4901	0,4138	0,2952
60	0,9784	0,8025	0,7086	0,6472	0,6028	0,5687	0,5189	0,4574	0,3746	0,2352
120	0,9622	0,7829	0,6867	0,6234	0,5774	0,5419	0,4897	0,4243	0,3339	0,1612
∞	0,9462	0,7636	0,6651	0,5999	0,5522	0,5152	0,4604	0,3908	0,2913	0,0000

Tafel VIII. *Die Kolmogoroff-Smirnowsche λ-Verteilung*

Die Tafel gibt Werte der Funktion $Q(\lambda)$ an, die durch die Formel

$$Q(\lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}$$

bestimmt ist.

λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$
0,32	0,0000	0,66	0,2236	1,00	0,7300	1,34	0,9449	1,68	0,9929	2,00	0,9993
0,33	0,0001	0,67	0,2396	1,01	0,7406	1,35	0,9478	1,69	0,9934	2,01	0,9994
0,34	0,0002	0,68	0,2558	1,02	0,7508	1,36	0,9505	1,70	0,9938	2,02	0,9994
0,35	0,0003	0,69	0,2722	1,03	0,7608	1,37	0,9531	1,71	0,9942	2,03	0,9995
0,36	0,0005	0,70	0,2888	1,04	0,7704	1,38	0,9556	1,72	0,9946	2,04	0,9995
0,37	0,0008	0,71	0,3055	1,05	0,7798	1,39	0,9580	1,73	0,9950	2,05	0,9996
0,38	0,0013	0,72	0,3223	1,06	0,7889	1,40	0,9603	1,74	0,9953	2,06	0,9996
0,39	0,0019	0,73	0,3391	1,07	0,7976	1,41	0,9625	1,75	0,9956	2,07	0,9996
0,40	0,0028	0,74	0,3560	1,08	0,8061	1,42	0,9646	1,76	0,9959	2,08	0,9996
0,41	0,0040	0,75	0,3728	1,09	0,8143	1,43	0,9665	1,77	0,9962	2,09	0,9997
0,42	0,0055	0,76	0,3896	1,10	0,8223	1,44	0,9684	1,78	0,9965	2,10	0,9997
0,43	0,0074	0,77	0,4064	1,11	0,8299	1,45	0,9702	1,79	0,9967	2,11	0,9997
0,44	0,0097	0,78	0,4230	1,12	0,8374	1,46	0,9718	1,80	0,9969	2,12	0,9997
0,45	0,0126	0,79	0,4395	1,13	0,8445	1,47	0,9734	1,81	0,9971	2,13	0,9998
0,46	0,0160	0,80	0,4559	1,14	0,8514	1,48	0,9750	1,82	0,9973	2,14	0,9998
0,47	0,0200	0,81	0,4720	1,15	0,8580	1,49	0,9764	1,83	0,9975	2,15	0,9998
0,48	0,0247	0,82	0,4880	1,16	0,8644	1,50	0,9778	1,84	0,9977	2,16	0,9998
0,49	0,0300	0,83	0,5038	1,17	0,8706	1,51	0,9791	1,85	0,9979	2,17	0,9998
0,50	0,0361	0,84	0,5194	1,18	0,8765	1,52	0,9803	1,86	0,9980	2,18	0,9999
0,51	0,0428	0,85	0,5347	1,19	0,8823	1,53	0,9815	1,87	0,9981	2,19	0,9999
0,52	0,0503	0,86	0,5497	1,20	0,8877	1,54	0,9826	1,88	0,9983	2,20	0,9999
0,53	0,0585	0,87	0,5645	1,21	0,8930	1,55	0,9836	1,89	0,9984	2,21	0,9999
0,54	0,0675	0,88	0,5791	1,22	0,8981	1,56	0,9846	1,90	0,9985	2,22	0,9999
0,55	0,0772	0,89	0,5933	1,23	0,9030	1,57	0,9855	1,91	0,9986	2,23	0,9999
0,56	0,0876	0,90	0,6073	1,24	0,9076	1,58	0,9864	1,92	0,9987	2,24	0,9999
0,57	0,0987	0,91	0,6209	1,25	0,9121	1,59	0,9873	1,93	0,9988	2,25	0,9999
0,58	0,1104	0,92	0,6343	1,26	0,9164	1,60	0,9880	1,94	0,9989	2,26	0,9999
0,59	0,1228	0,93	0,6473	1,27	0,9206	1,61	0,9888	1,95	0,9990	2,27	0,9999
0,60	0,1357	0,94	0,6601	1,28	0,9245	1,62	0,9895	1,96	0,9991	2,28	0,9999
0,61	0,1492	0,95	0,6725	1,29	0,9283	1,63	0,9902	1,97	0,9991	2,29	0,9999
0,62	0,1632	0,96	0,6846	1,30	0,9319	1,64	0,9908	1,98	0,9992	2,30	0,9999
0,63	0,1778	0,97	0,6964	1,31	0,9354	1,65	0,9914	1,99	0.9993	2,31	1,0000
0,64	0,1927	0,98	0,7079	1,32	0,9387	1,66	0,9919				
0,65	0,2080	0,99	0,7191	1,33	0,9418	1,67	0,9924				

LITERATUR

AHRENS, H.

[1] Varianzanalyse. 2. Aufl., Akademie-Verlag, Berlin 1968.

ANDERSON, T. W.

[1] Introduction to multivariate statistical analysis. John Wiley, New York—London 1958.

[2] A modification of the sequential probability ratio test to reduce the sample size. Annals of Math. Statistics **31** (1960), 165.

[3] Least squares and best unbiased estimates. Annals of Math. Statistics **33** (1962), 266.

ANDERSON, T. W., and D. A. DARLING

[1] Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Annals of Math. Statistics **23** (1952), 193.

ANDERSON, T. W., and L. A. GOODMAN

[1] Statistical inference about chains. Annals of Math. Statistics **28** (1957), 89.

ANSCOMBE, F. J.

[1] Large sample theory of sequential estimation. Proc. Cambridge Philos. Soc. **48** (1952), 600.

ARLEY, N.

[1] On the theory of stochastic processes and their application to the theory of cosmic radiation. John Wiley, New York 1948.

ARMITAGE, P.

[1] A comparison of stratified and unrestricted random sampling from a finite population. Biometrika **34** (1947), 273.

[2] Restricted sequential procedures. Biometrika **44** (1957), 9.

AUSTIN, D. G.

[1] On the existence of the derivate of Markoff transition probability functions. Proc. Nat. Acad. Sci. USA **41** (1955), 224.

[2] Note on differentiating Markoff transition functions with stable terminal states. Duke Math. J. **25** (1958), 625.

BABKIN, W. I., P. F. BELJAJEW und J. I. MAXIMOW (Бабкин, В. И., П. Ф. Беляев и Ю. И. Максимов)

[1] Некоторые замечания к работе В. Л. Гончарова «Из области комбинаторики». Теория вероятностей и ее применения **4** (1959), 445.

BACHELIER, L.

[1] Théorie de la spéculation. Ann. École Normale Sup. Paris **17** (1900), 21.

BARANKIN, E. W.

[1] Extension of a theorem of Blackwell. Annals of Math. Statistics **21** (1950), 280.

[2] Toward an objectivistic theory of probability. Proc. Third Berkeley Symp. **5** (1956), 21.

BARANKIN, E. W., and M. KATZ jr.

[1] Sufficient statistics of minimal dimension. Sankhya **21** (1959), 217.

BARNARD, G. A.

[1] On the Fisher-Behrens test. Biometrika **37** (1950), 203.

BARTKY, W.

[1] Multiple sampling with constant probability. Annals of Math. Statistics **14** (1943), 363.

BARTLETT, M. S.

[1] On the theory of statistical regression. Proc. Royal Soc. Edinburgh **53** (1933), 260.

BARTON, D. E., and F. N. DAVID

[1] Non-randomness in a sequence of two alternatives, II. Run tests. Biometrika **45** (1958), 253.

BARTON, D. E., F. N. DAVID and C. L. MALLOWS

[1] Non-randomness in a sequence of two alternatives, I. Wilcoxon's and allied test statistics. Biometrika **45** (1958), 166.

BARTOSZYŃSKI, R.

[1] Some remarks on the convergence of stochastic processes. Studia Math. **17** (1958), 313.

[2] A characterization of the weak convergence of measures. Annals of Math. Statistics **32** (1961), 561.

BASU, D., and R. G. LAHA

[1] On some characterizations of the normal distribution. Sankhya **13** (1954), 359.

BAXTER, G., and M. D. DONSKER

[1] On the distribution of the supremum functional for processes with stationary independent increments. Trans. Amer. Math. Soc. **85** (1957), 73.

BAŻAŃSKA, T.

[1] Zastosowanie metod statystyki matematycznej do obliczenia mocy heparyny (unveröffentlicht).

BELJAJEW, J. K. (Беляев, Ю. К.)

[1] О неограниченности выборочных функций гауссовских процессов. Теория вероятностей и ее применения **3** (1958), 351.

[2] Аналитические случайные процессы. Теория вероятностей и ее применения **4** (1959), 437.

[3] Локальные свойства выборочных функций стационарных гауссовских процессов. Теория вероятностей и ее применения **5** (1960), 128.

BELL, C. B.

[1] On the structure of distribution-free statistics. Annals of Math. Statistics **31** (1960), 703.

BERGE, P. O.

[1] A note on a form of Tschebycheff's theorem for two variables. Biometrika **29** (1957), 405.

BERNSTEIN, S. N. (Бернштейн, С. Н.)

[1] Опыт аксиоматического обоснования теории вероятностей. Харк. Зап. Матем. о-ва **15** (2) (1917), 209.

[2] Sur les sommes de quantités dépendantes. Известия Акад. Наук **20** (6) (1926), 1459.

[3] Sur l'extension du théorème limite du calcul des probabilités aux sommes des quantités dépendantes. Math. Annalen **97** (1926), 1.

[4] Об одном свойстве, характеризующем закон Гаусса. Труды Л. Политехн. ин-та **3** (1941), 3.

BERRY, A. C.

[1] The accuracy of the Gaussian approximation of the sum of independent variates. Trans. Amer. Math. Soc. **49** (1941), 122.

BHABHA, H. J., and W. HEITLER

[1] The passage of fast electrons and the theory of cosmic showers. Proc. Royal Soc. London A **159** (1937), 432.

BHARUCHA-REID, A. T.

[1] Note on estimation of the number of states in a discrete Markov chain. Experientia **12** (1956), 176.

[2] Elements of the theory of Markov processes and their applications. McGraw-Hill, New York—Toronto—London 1960.

BILLINGSLEY, P.

[1] The invariance principle for dependent random variables. Trans. Amer. Math. Soc. **83** (1956), 250.

[2] Statistical methods in Markov chains. Annals of Math. Statistics **32** (1961), 13.

[3] The Lindeberg-Lévy theorem for martingales. Proc. Amer. Math. Soc. **12** (1961), 788.

BIRKHOFF, G. D.

[1] Proof of the ergodic theorem. Proc. Nat. Acad. Sci. USA **17** (1931), 656.

BIRNBAUM, A.

[1] A unified theory of estimation. I. Annals of Math. Statistics **32** (1962), 112.

[2] On the foundations of statistical inference (with discussion). J. Amer. Stat. Assn. **57** (1962), 269.

BIRNBAUM, Z. W.

[1] Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. J. Amer. Stat. Assn. **47** (1952), 425.

[2] Distribution-free tests of fit for continuous distribution functions. Annals of Math. Statistics **24** (1953), 1.

[3] On the power of a one-sided test of fit for continuous probability functions. Annals of Math. Statistics **24** (1953), 484.

[4] On a use of the Mann-Whitney statistic. Proc. Third Berkeley Symp. **1** (1955), 13.

BIRNBAUM, Z. W., and R. A. HALL

[1] Small sample distributions for multi-sample statistics of the Smirnov type. Annals of Math. Statistics **31** (1960), 710.

BIRNBAUM, Z. W., and A. W. MARSHALL

[1] Some multivariate Chebyshev inequalities with extension to continuous parameter processes. Annals of Math. Statistics **32** (1961), 687.

BIRNBAUM, Z. W., and R. C. MCCARTY

[1] A distribution-free upper confidence bound for $Pr\{Y < X\}$ based on independent samples of X and Y. Annals of Math. Statistics **29** (1958), 558.

BIRNBAUM, Z. W., and R. PYKE

[1] On some distributions related to the statistic D_n^+. Annals of Math. Statistics **29** (1958), 179.

BIRNBAUNM, Z. W., and H. RUBIN

[1] On distribution-free statistics. Annals of Math. Statistics **25** (1954), 593.

BIRNBAUM, Z. W., and F. H. TINGEY

[1] One-sided confidence contours for probability distribution functions. Annals of Math. Statistics **22** (1951), 592.

BLACKWELL, D.

[1] On an equation of Wald. Annals of Math. Statistics **17** (1946), 84.

[2] Conditional expectation and unbiased sequential estimation. Annals of Math. Statistics **18** (1947), 105.

BLACKWELL, D., and M. A. GIRSHICK
[1] On functions of sequences of independent chance vectors with applications to the problem of the "random walk" in k dimensions. Annals of Math. Statistics **17** (1946), 310.
[2] A lower bound for the variance of some unbiased sequential estimates. Annals of Math. Statistics **18** (1947), 277.
BLANC-LAPIERRE, A., et R. FORTET
[1] Théorie des fonctions aléatoires. Masson, Paris 1953.
BLOM, G.
[1] A generalization of Wald's fundamental identity. Annals of Math. Statistics **20** (1949), 439.
BLOMBERG, C.
[1] Statistical analysis of prognoses. Kgl. tekn. högskol. handl., No. 203, Göteborg 1963.
BLUM, J. R.
[1] On the convergence of empiric distribution functions. Annals of Math. Statistics **26** (1955), 527.
BLUMENTHAL, R. M.
[1] An extended Markov property. Trans. Amer. Math. Soc. **85** (1957), 52.
BLUMENTHAL. R. M., and R. K. GETOOR
[1] Some theorems on stable processes. Trans. Amer. Math. Soc. **95** (1960), 263.
[2] A dimension theorem for sample functions of stable processes. Ill. J. Math. **4** (1960), 370.
BOBROFF, A. A. (BOBROV, A. A.; Бобров, A. A.)
[1] Conditions of applicability of the strong law of large numbers. Duke Math. J. **12** (1945), 43.
BOCHNER, S.
[1] Vorlesungen über Fouriersche Integrale. Akadem. Verlagsgesellschaft, Leipzig 1932.
[2] Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse. Math. Annalen **108** (1933), 378.
BOJEW, G. P. (Боев, Г. П.)
[1] Теория вероятностей. Гостехиздат, Москва 1950.
BOREL, E.
[1] Les probabilités dénombrables et leurs applications arithmétiques. Rend. Circ. Mat. Palermo **27** (1909), 247.
[2] Sur un problème de probabilités rélatif aux fractions continues. Math. Annalen **72** (1912), 578.
BORTKIEWICZ, L.
[1] Das Gesetz der kleinen Zahlen. Teubner, Leipzig 1898.
[2] Die Iterationen. Ein Beitrag zur Wahrscheinlichkeitsrechnung. Springer, Berlin 1917.
BOWKER, A. H.
[1] Tolerance limits for normal distributions. Techniques of Statistical Analysis, Statistical Research Group, Columbia University, New York 1947, p. 95.
BRADLEY, J. V.
[1] Distribution-free statistical tests. Prentice-Hall, Englewood Cliffs, N.J., 1968.
BREIMAN, L.
[1] A counterexample to a theorem of Kolmogorov. Annals of Math. Statistics **28** (1957), 811.
[2] The strong law of large numbers for a class of Markov chains. Annals of Math. Statistics **31** (1960), 801.
BRUNK, H. D.
[1] The strong law of great numbers. Duke Math. J. **15** (1948), 181.

BURKHARDT, F.

[1] Anwendungen der Wahrscheinlichkeitsrechnung und der mathematischen Statistik in der Wirtschaft, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin **1956**, S. 65.

CANTELLI, F. P.

[1] Sulla probabilità come limite della frequenza. Rend. Accad. Lincei **26** (1) (1917), 39.

CHANG, L. C.

[1] On the ratio of an empirical distribution function to the theoretical distribution function. Acta Math. Sinica **5** (1955), 347.

CHANG, L. C., and M. FISZ

[1] Asymptotically independent linear functions of empirical distribution functions. Sci. Record **1** (1957), 335.

[2] Exact distributions of the maximal values of some functions of empirical distribution functions. Sci. Record **1** (1957), 341.

CHAPMAN, D. G.

[1] Some two-sample tests. Annals of Math. Statistics **21** (1950), 601.

[2] Estimating the parameters of a truncated gamma distribution. Annals of Math. Statistics **27** (1956), 498.

[3] A comparative study of several one-sided goodness-of-fit tests. Annals of Math. Statistics **29** (1958), 655.

CHERNOFF, H., and E. L. LEHMANN

[1] The use of the maximum likelihood estimates in χ^2 test for goodness of fit. Annals of Math. Statistics **25** (1954), 579.

CHINTSCHIN, A. J. (KHINTCHINE, A.; CHINČIN, A. JA.; ХИНЧИН, А. Я.)

[1] Sur la loi des grands nombres. Comptes Rendus Paris **188** (1929), 477.

[2] Asymptotische Gesetze der Wahrscheinlichkeitsrechnung. Springer, Berlin 1933.

[3] Korrelationstheorie der stationären stochastischen Prozesse. Math. Annalen **109** (1934), 604.

[4] Sul dominio di attrazione della legge di Gauss. Giorn. Ist. Ital. Attuari **6** (1935), 371.

[5] Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. Матем. Сб. **2** (44) (1937), 79.

[6] Предельные законы для сумм независимых случайных величин. ОНТИ, Москва—Ленинград 1938.

[7] Математические методы теории массового обслуживания. Труды Матем. ин-та АН СССР **49** (1955).

[8] Über einen Satz der Wahrscheinlichkeitsrechnung. Fundamenta Math. **6** (1924), 9.

CHINTSCHIN, A. J., und A. N. KOLMOGOROFF (Хинчин, А. Я., и А. Н. Колмогоров)

[1] Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. Матем. Сб. **32** (1925), 668.

CHOW, Y. S.

[1] Martingales in a σ-finite measure space indexed by directed sets. Trans. Amer. Math. Soc. **97** (1960), 254.

CHUNG, K. L.

[1] The strong law of large numbers. Proc. Second Berkeley Symp. 1951, p. 341.

[2] Contributions to the theory of Markov chains I. J. Res. Nat. Bureau Standards **50** (1953), 203.

[3] Contributions to the theory of Markov chains. Trans. Amer. Math. Soc. **76** (1954), 397.

[4] Some new developments in Markov chains. Trans. Amer. Math. Soc. **81** (1956), 195.
[5] Markov chains with stationary transition probabilities. Springer-Verlag, Berlin—Göttingen—Heidelberg 1960.

Chung, K. L., and W. Feller
[1] On fluctuations of coin tossing. Proc. Nat. Acad. Sci. USA **35** (1949), 605.

Ciesielski, Z.
[1] Hölder conditions for realizations of Gaussian processes. Trans. Amer. Math. Soc. **99** (1961), 403.

Cochran, W. G.
[1] Relative accuracy of systematic and stratified samples for a certain class of populations. Annals of Math. Statistics **17** (1946), 164.
[2] The χ^2-test of goodness of fit. Annals of Math. Statistics **23** (1952), 315.
[3] Sampling techniques, 2nd ed., John Wiley & Sons, New York—London 1963.

Cochran, W. G., and G. M. Cox
[1] Experimental designs. John Wiley & Sons, New York 1957.

Cogburn, R.
[1] Lois limites des termes variationnels des sommes normées. Comptes Rendus Paris **246** (1958), 3408.

Courant, R., und D. Hilbert
[1] Methoden der Mathematischen Physik. Springer, Göttingen 1930.

Cowden, D. J.
[1] Statistical methods in quality control. Prentice-Hall, Englewood Cliffs, N.J., 1957.

Cramér, H.
[1] Random variables and probability distributions. University Press, Cambridge 1937.
[2] Mathematical methods in statistics. Princeton University Press, Princeton 1946.
[3] On the composition of elementary errors. Skand. Aktuar. **11**, Nr. 13 (1928), 141.
[4] Über eine Eigenschaft der normalen Verteilungsfunktion. Math. Z. **41** (1936), 405.
[5] On the representation of a function by certain Fourier integrals. Trans. Amer. Math. Soc. **46** (1939), 191.
[6] A contribution to the theory of stochastic processes. Proc. Second Berkeley Symp. 1951, p. 329.

Cramér, H., and H. Wold
[1] Some theorems on distributions. J. London Math. Soc. **11** (1936), 290.

Crow, E. L., F. A. Davis and M. W. Maxfield
[1] Statistics manual. Dover Publ., Inc., New York 1960.

Czencow, N. N. (Čenzov, N. N.; Tschenzow, N. N.; Ченцов, Н. Н.)
[1] Слабая сходимость случайных процессов с траекториями без разрывов второго рода и так называемый ,,эвристический'' подход к критериям согласия типа Колмогорова-Смирнова. Теория вероятностей и ее применения **1** (1956), 155.

Czuber, E.
[1] Die philosophischen Grundlagen der Wahrscheinlichkeitstheorie. Teubner, Leipzig - Berlin 1923.

Dalenius, T.
[1] The multivariate sampling problem. Skand. Aktuar. **36** (1953), 92.
[2] Sampling in Sweden. Almqvist and Wiksell, Stockholm 1957.

DANIELS, H. E.

[1] Some problems of statistical interest in wool research. Supplement J. Royal Statist. Soc. **5** (1938), 89.

[2] The statistical theory of the strength of bundles of threads I. Proc. Royal Soc. London A **189** (1945), 405.

VAN DANTZIG, D.

[1] On the consistency and the power of Wilcoxon's two sample test. Proc. Nederl. Akad. Wet. Amsterdam A **54** (1951), 1.

DANTZIG, G. B.

[1] On the non-existence of tests of "Student's hypothesis" having power functions independent of σ. Annals of Math. Statistics **11** (1940), 186.

DANTZIG, G. B., and A. WALD

[1] On the fundamental lemma of Neyman and Pearson. Annals of Math. Statistics **22** (1951), 87.

DARLING, D. A.

[1] The Kolmogorov-Smirnov, Cramér-von Mises tests. Annals of Math. Statistics **28** (1957), 823.

DARMOIS, G.

[1] Sur les lois des probabilités à estimation exhaustive. Comptes Rendus Paris **200** (1935), 1265.

[2] Sur une propriété caractéristique de la loi de probabilité de Laplace. Comptes Rendus Paris **232** (1951), 1999.

DAVID, F. N.

[1] Tables of the correlation coefficient. London 1938.

DAVID, F. N., and J. NEYMAN

[1] Extension of the Markoff theorem on least squares. Stat. Res. Mem. **2** (1938), 105.

DAVID, H. T.

[1] A three-sample Kolmogorov-Smirnov test. Annals of Math. Statistics **29** (1958), 842.

DAVIES, O. L., and E. S. PEARSON

[1] Methods of estimating from samples the population standard deviation. Supplement J. Royal Statist. Soc. **1** (1934), 76.

DEMING, W. E.

[1] Some theory of sampling. John Wiley, New York 1950.

DEMPSTER, A. P.

[1] Generalized D_n^+ statistics. Annals of Math. Statistics **30** (1959), 593.

DERMAN, C., S. LITTAUER and H. SOLOMON

[1] Tightened multi-level continuous sampling plans. Annals of Math. Statistics **28** (1957), 395.

DOBRUSCHIN, R. L. (Добрушин, Р. Л.)

[1] Центральная предельная теорема для неоднородных цепей Маркова. Доклады Акад. Наук СССР **102** (1955), 5.

[2] Об условиях регулярности однородных по времени марковских процессов со счетным числом возможных состояний. Успехи Матем. Наук **7**: 6 (52) (1952), 185.

[3] Условия регулярности марковских процессов с конечным числом возможных состояний. Матем. Сб. **34** (76) (1954), 541.

[4] Лемма о пределе сложной случайной функции. Успехи Матем. Наук **10**: 2 (64) (1955), 157.

[5] Условие непрерывности выборочных функций мартингала. Теория вероятностей и ее применения **3** (1958), 97.
[6] Свойства выборочных функций гауссовского стационарного процесса. Теория вероятностей и ее применения **5** (1960), 132.

Dodge, H. F., and H. G. Romig
[1] A method of sampling inspection. Bell system Tech. J. 8 (1929), 613.

Doeblin, W.
[1] Sur les propriétés asymptotiques de mouvements régis par certains types des chaînes simples. Bull. Math. Soc. Roum. Sci. **39** (2) (1937), 3.
[2] Sur deux problèmes de M. Kolmogoroff concernant les chaînes dénombrables. Bull. Math. Soc. France **66** (1938), 210.
[3] Sur certains mouvements aléatoires discontinus. Skand. Aktuar. **22** (1939), 21.
[4] Éléments d'une théorie générale des chaînes simples constantes de Markoff. Ann. École Normale Sup. Paris (3) **57** (1940), 61.

Donsker, M. D.
[1] An invariance principle for certain probability limit theorems. Memoirs Amer. Math. Soc. **6** (1951), 1.
[2] Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov limit theorems. Annals of Math. Statistics **23** (1952), 277.

Doob, J. L.
[1] Stochastic processes depending on a continuous parameter. Trans. Amer. Math. Soc. **42** (1937), 107.
[2] Topics in the theory of Markoff chains. Trans. Amer. Math. Soc. **52** (1942), 37.
[3] Markoff chains—denumerable case. Trans. Amer. Math. Soc. **58** (1945), 455.
[4] Heuristic approach to the Kolmogorov-Smirnov theorems. Annals of Math. Statistics **20** (1949), 393.
[5] Stochastic processes. John Wiley, New York 1953.
[6] Probability and statistics. Trans. Amer. Math. Soc. **36** (1934), 759.
[7] Regularity properties of certain families of chance variables. Trans. Amer. Math. Soc. **47** (1940), 455.
[8] Application of the theory of martingales. Le Calcul des Probabilités et ses applications. Coll. Intern. Center Nat. Rep. Sci., Paris **23** (1949).
[9] Continuous parameter martingales. Proc. Second Berkeley Symp. 1951, p. 269.
[10] Discrete potential theory and boundaries. J. Math. Mech. **8** (1959), 433.
[11] Notes on martingale theory. Proc. Fourth Berkeley Symp. **2** (1961), 95.

Dugué, D.
[1] Application des propriétés de la limite au sens du calcul des probabilités à l'étude des diverses questions d'estimation. J. École Polytechnique 1937, p. 305.
[2] Traité de statistique théorique et appliquée. Analyse aléatoire—algèbre aléatoire. Masson, Paris 1958.

Duncan, A. J.
[1] Quality control and industrial statistics. Homewood 1959.

Dupač, V., a J. Hájek
[1] Pravděpodobnost ve vědě a technice. Nakladatelstvi Československé akademie věd., Praha 1962.

Dvoretsky, A.
[1] On the strong stability of a sequence of events. Annals of Math. Statistics **20** (1949), 296.

DVORETSKY, A., P. ERDÖS and S. KAKUTANI

[1] Nonincrease everywhere of the Brownian Motion Process. Proc. Fourth Berkeley Symp. **2** (1961), 103.

DVORETSKY, A., J. KIEFER and J. WOLFOWITZ

[1] Sequential decision problems for processes with continuous time parameter. Testing hypotheses. Annals of Math. Statistics **24** (1953), 254.

[2] Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Annals of Math. Statistics **27** (1956), 642.

DVORETSKY, A., and J. WOLFOWITZ

[1] Sums of random integers reduced modulo m. Duke Math. J. **18** (1951), 501.

DWASS, M.

[1] The distribution of a generalized D_n^+ statistic. Annals of Math. Statistics **30** (1959), 1024.

[2] Some k-sample rank order tests. Contributions to probability and statistics. Essays in honor of Harold Hotelling, Stanford University Press 1960, p. 504.

DYNKIN, E. B. (ДЫНКИН, Е. Б.)

[1] Критерии непрерывности и отсутствия разрывов второго рода для траекторий марковского случайного процесса. Известия Акад. Наук УССР **16** (1952), 573.

[2] О некоторых предельных теоремах для цепей Маркова. Укр. Матем. Ж. **6** (1954), 21.

DYNKIN, E. B., und W. A. USPENSKI

[1] Mathematische Unterhaltungen, Teil III, 3. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1967 (Übersetzung aus dem Russischen).

EHRENFEST, P. and T.

[1] Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. Phys. Z. **8** (1907), 311.

EHRENFEUCHT, A., and M. FISZ

[1] A necessary and sufficient condition for the validity of the weak law of large numbers. Bull. Pol. Acad. Sci. **8** (1960), 583.

EINSTEIN, A.

[1] Zur Theorie der Brownschen Bewegung. Ann. Phys. IV **19** (1906), 371.

ELFVING, E. G.

[1] The asymptotical distribution of range in samples from a normal population. Biometrika **35** (1947), 111.

[2] Über optimale Allokation, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 89.

ERDÖS, P.

[1] On the law of the iterated logarithm. Annals Math. (2) **43** (1942), 419.

ERDÖS, P., and M. KAC

[1] On certain limit theorems of the theory of probability. Bull. Amer. Math. Soc. **53** (1946), 292.

[2] On the number of positive sums of independent random variables. Bull. Amer. Math. Soc. **53** (1947), 1011.

ERDÖS, P., and A. RÉNYI

[1] On the central limit theorem for samples from a finite population. Matem. Kutato Intezet Közlem. **4** (1959), 49.

[2] On Cantor's series with convergent $\sum 1/q^n$. Ann. Univ. Sci. Budapestinensis, Sectio Math., **2** (1959), 93.

ERLANG, A. K.

[1] The life and work of A. K. Erlang. The Copenhagen Telephone Co. 1948.

ESSEEN, C. G.

[1] Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. Acta Math. **77** (1945), 1.

[2] A moment inequality with an application to the central limit theorem. Skand. Aktuar. **39** (1956), 160.

EZEKIEL, M.

[1] Methods of correlation analysis. John Wiley, New York 1950.

FABIAN, V.

[1] Statistische Methoden. 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1970 (Übersetzung aus dem Tschechischen).

FELLER, W.

[1] Zur Theorie der stochastischen Prozesse. Math. Annalen **113** (1936), 113.

[2] Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. Math. Z. **40** (1935), 521; **42** (1937), 301.

[3] Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. Acta Biotheoretica **5** (1939), 11.

[4] On the integro-differential equations of purely discontinuous Markov processes. Trans. Amer. Math. Soc. **48** (1940), 488.

[5] On the Kolmogorov-Smirnov limit theorems for empirical distributions. Annals of Math. Statistics **19** (1948), 177.

[6] On the theory of stochastic processes, with particular reference to applications. Proc. Berkeley Symp. 1949, p. 403.

[7] An introduction to probability theory and its applications. John Wiley, New York 1950.

[8] Über das Gesetz der Großen Zahlen. Acta Litt. Acad. Sci. Sect. Math. Szeged **8** (1936/37), 191.

[9] Note on regions similar to the sample space. Stat. Res. Mem. **2** (1938), 117.

[10] The general form of the so-called law of the iterated logarithm. Trans. Amer. Math. Soc. **54** (1943), 373.

[11] On the normal approximation to the binomial distribution. Annals of Math. Statistics **16** (1945), 319.

FERGUSON, T.

[1] A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. Annals of Math. Statistics **29** (1958), 1046.

[2] Mathematical statistics — A decision theoretic approach. 3. ed., Academic Press, New York 1969.

FERRIS, C. D., F. E. GRUBBS and C. L. WEAVER

[1] Operating characteristics for the common statistical tests of significance. Annals of Math. Statistics **17** (1946), 178.

FICHTENHOLZ, G. M. (Фихтенгольц, Г. М.)

[1] Differential- und Integralrechnung. Band II, 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1966 (Übersetzung aus dem Russischen).

DE FINETTI, B.

[1] Sulle funzioni a incremento aleatorio. Rend. Accad. Lincei Cl. Sci. Fis. Mat. **10** (6) (1929), 163.

FISHER, R. A.

[1] Frequency distributions of the values of the correlation coefficient in samples from an infinitely large population. Biometrika **10** (1915), 507.

[2] On the mathematical foundations of theoretical statistics. Phil. Trans. Royal Soc. **222** (1921), 309.

[3] On the "probable error" of a coefficient of correlation deduced from a small sample. Metron **1**, Nr. 4 (1921), 1.

[4] The conditions under which χ^2 measures the discrepancy between observation and hypothesis. J. Royal Statist. Soc. **87** (1924), 442.

[5] Application of Student distribution. Metron **5**, Nr. 3 (1925), 90.

[6] Theory of statistical estimation. Proc. Cambridge Philos. Soc. **22** (1925), 700.

[7] Statistical methods for research workers. Oliver and Boyd, Edinburgh–London 1941.

[8] On an absolute criterion for fitting frequency curves. Messenger of Math. **41** (1912), 155.

[9] On a property connecting the χ^2 measure of discrepancy with the method of maximum likelihood. Atti Congr. Intern. Mat. Bologna **6** (1928), 94.

[10] Two new properties of mathematical likelihood. Proc. Royal Soc. London A **144** (1934), 285.

[11] The logic of inductive inference. J. Royal Statist. Soc. **98** (1935), 39.

[12] The fiducial argument in statistical inference. Ann. Eugenics Cambridge **6** (1936), 391.

[13] The distribution of the partial correlation coefficient. Metron **3** (1924), 329.

[14] On a distribution yielding the error functions of several well-known statistics. Proc. Intern. Math. Congr. Toronto 1924, p. 805.

[15] The general sampling distribution of the multiple correlation coefficient. Proc. Royal Soc. London A **121** (1928), 654.

FISHER, R. A., and F. YATES

[1] Statistical tables. Oliver and Boyd, London 1938.

[2] Statistical tables for biological, agricultural and medical research. London–Edinburgh 1957.

FISZ, M.

[1] The limiting distribution of sums of arbitrary independent and equally distributed r-point random variables. Studia Math. **14** (1953), 111.

[2] Die Grenzverteilungen der Multinomialverteilung, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 51.

[3] Realizations of some stochastic processes. Studia Math. **15** (1956), 359.

[4] A limit theorem for empirical distribution functions. Bull. Acad. Polon. Sci. **5** (1957), 699; Studia Math. **17** (1958), 71.

[5] A limit theorem for non-decreasing random functions. Bull. Acad. Polon. Sci. **6** (1958), 485.

[6] On necessary and sufficient conditions for the validity of the strong law of large numbers expressed in terms of moments. Bull. Acad. Polon. Sci. **7** (1959), 229.

[7] Characterization of sample functions of stochastic processes by some absolute probabilities. Proc. Fourth Berkeley Symp. **2** (1961), 143.

FISZ, M., and K. URBANIK

[1] Analytical characterization of some composed non-homogeneous Poisson process. Studia Math. **15** (1956), 328.

FIX, E.

[1] Tables of the noncentral χ^2. Univ. Calif. Publ. Stat. **1** (1949), 15.

FIX, E., and J. L. HODGES

[1] Significance probabilities of the Wilcoxon test. Annals of Math. Statistics **26** (1955), 301.

FLACHSMEYER, J.

[1] Kombinatorik. Eine Einführung in die mengentheoretische Denkweise. 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1970.

Florek, K., E. Marczewski and C. Ryll-Nardzewski
[1] Remarks on the Poisson stochastic process I. Studia Math. **13** (1953), 122.
Fogelson, S.
[1] Mediana i jej wyznaczenie. Kwartalnik Statystyczny **7** (1930) 867.
Fortet, R.
[1] Quelques travaux récents sur le mouvement Brownien. Ann. Inst. H. Poincaré **11** (1949), 175.
[2] Calcul des probabilités. Centre Nat. Res. Sci., Paris 1950.
[3] Random functions from a Poisson process. Proc. Second Berkeley Symp. 1951, p. 373.
[4] Normalverteilte Zufallselemente in Banachschen Räumen. Anwendungen auf zufällige Funktionen, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 29.
Fortet, R., et E. Mourier
[1] Convergence de la répartition empirique vers la répartition théorique. Comptes Rendus Paris **236** (1953), 1739.
Fraser, D. A. S.
[1] Nonparametric methods in statistics. John Wiley, New York 1957.
Fréchet, M.
[1] Recherches théoretiques modernes sur la théorie des probabilités. II. Gauthier-Villars, Paris 1937/38.
[2] Sur la loi de probabilité de l'écart maximum. Ann. Soc. Polon. Math. **6** (1927), 93.
[3] Abstrakte Zufallselemente, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 23.
Fréchet, M., and J. Shohat
[1] A proof of the generalized second limit theorem in the theory of probability. Trans. Amer. Math. Soc. **33** (1931), 533.
Freudenthal, H.
[1] Wahrscheinlichkeit und Statistik. Oldenbourg, München 1963.
Frisch, R.
[1] Correlation and scatter in statistical variables. Nord. Stat. Tidskr. **8** (1929), 36.
[2] Statistical confluence analysis by means of complete regression systems. Oslo 1934.
Fryde, M.
[1] Teoria i rzeczywistość w statystyce. Economista **3** (1923), 28.
Fuchs, A.
[1] Sur la continuité stochastique des processus stochastiques réels de Markoff. Comptes Rendus Paris **237** (1953), 1329.
[2] Some limit theorems for non-homogeneous Markov processes. Trans. Amer. Math. Soc. **86** (1957), 511.
Furry, W. H.
[1] On fluctuation phenomena in the passage of high energy electron through lead. Phys. Rev. **52** (1937), 569.
Gabriel, K. R.
[1] The distribution of the number of successes in a sequence of dependent trials. Biometrika **46** (1959), 454.
Gantmacher, F. R. (Гантмахер, Ф. Р.)
[1] Matrizenrechnung, Band I, 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1965 (Übersetzung aus dem Russischen).

GAUSS, C. F.

[1] Teoria motus corporum coelestium. Perthes and Besser, Hamburg 1809.

GEARY, R. C.

[1] Distribution of Student's ratio for non-normal samples. Supplement J. Royal Statist. Soc. **3** (1936), 178.

[2] The estimation of many parameters. J. Royal Statist. Soc. **105** (1942), 213.

GEBELEIN, H.

[1] Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. Z. angew. Math. Mech. **21** (1941), 364.

GEISSER, S., and N. MANTEL

[1] Pairwise independence of jointly dependent variables. Annals of Math. Statistics **33** (1962), 290.

GETOOR, R. K.

[1] The shift operator for non-stationary stochastic processes. Duke Math. J. **23** (1956), 175.

[2] On characteristic functions of Banach space valued random variables. Pacific J. Math. **7** (1957), 885.

GHOSH, P. S.

[1] A note on stratified random sampling with multiple characteristics. Calcutta Stat. Assn. Bulletin **8** (1958), 81.

GICHMAN, I. I. (Гихман, И. И.)

[1] Об эмпирической функции распределения в случае группировки данных. Доклады Акад. Наук СССР **82** (1952), 837.

[2] О некоторых предельных теоремах для условных распределений и о связанных с ними задачах математической статистики. Укр. Матем. Ж. **5** (1953), 413.

[3] Процессы Маркова в задачах математической статистики. Укр. Матем. Ж. **6** (1954), 28.

[4] Об асимптотических свойствах некоторых статистик, аналогичных величине χ^2. Теория вероятностей и ее применения **1** (1956), 344.

[5] Об одном непараметрическом критерии однородности k выборок. Теория вероятностей и ее применения **2** (1957), 380.

GICHMAN, I. I., B. W. GNEDENKO und N. W. SMIRNOW (Гихман, И. И., Б. В. Гнеденко и Н. В. Смирнов)

[1] Непараметрические методы статистики. Труды 3-го Всес. Матем. съезда, Москва 1956, **3** (1958), 320.

GILBERT, W. M.

[1] Projections of probability distributions. Acta Math. Acad. Sci. Hung. **6** (1955), 195.

GIRAULT, M.

[1] Stochastic processes. Springer-Verlag, Berlin—Heidelberg—New York 1966.

GIRSHICK, M. A.

[1] Contributions to the theory of sequential analysis I—III. Annals of Math. Statistics **17** (1946), 123, 282.

GLIVENKO, V. I. (Гливенко, В. И.; GLIWENKO, W. I.)

[1] Sulla determinazione empirica delle leggi di probabilità. Giorn. Ist. Ital. Attuari **4** (1933), 92.

GNEDENKO, B. W. (Гнеденко, Б. В.)

[1] О характеристических функциях. Моск. Бюлл. ун-та (А) **1** (5) (1937), 17.

[2] О сходимости законов распределения сумм независимых слагаемых. Доклады Акад. Наук СССР **18** (1938), 231.

[3] К теории предельных теорем для сумм независимых случайных величин. Известия Акад. Наук **1939**, стр. 181, 643.

[4] Предельные законы для сумм независимых случайных величин. Успехи Матем. Наук **10** (1944), 115.

[5] О локальной предельной теореме теории вероятностей. Успехи Матем. Наук **3**: 3 (1948), 187.

[6] Lehrbuch der Wahrscheinlichkeitsrechnung. 5. Aufl., Akademie-Verlag, Berlin 1968 (Übersetzung aus dem Russischen).

[7] Проверка неизменности распределения вероятностей в двух независимых выборках. Math. Nachrichten **12** (1954), 29.

[8] Sur la distribution limite du terme maximum d'une série aléatoire. Annals Math. **44** (1943), 423.

[9] Элементы теории функций распределения случайных векторов. Успехи Матем. Наук **10** (1944), 230.

[10] Об одной теореме С. Н. Бернштейна. Известия Акад. Наук, сер. матем., **12** (1948), 97.

[11] О критерии Вилькоксона сравнения двух выборок. Bull. Acad. Polon. Sci. **6** (1958), 611.

[12] Über die Nachprüfung statistischer Hypothesen mit Hilfe der Variationsreihe, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 97.

Gnedenko, B. W., J. K. Beljajew und A. D. Solowjew (Гнеденко, Б. В., Ю. К. Беляев и А. Д. Соловьев)

[1] Mathematische Methoden der Zuverlässigkeitstheorie I, II. Akademie-Verlag, Berlin 1968 (Übersetzung aus dem Russischen).

Gnedenko, B. W., und A. J. Chintschin (Гнеденко, Б. В., и А. Я. Хинчин)

[1] Elementare Einführung in die Wahrscheinlichkeitsrechnung. 7. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1967 (Übersetzung aus dem Russischen).

Gnedenko, B. W., und A. N. Kolmogorov (Гнеденко, Б. В., и А. Н. Колмогоров)

[1] Grenzverteilungen von Summen unabhängiger Zufallsgrößen. 2. Aufl., Akademie-Verlag, Berlin 1960 (Übersetzung aus dem Russischen).

Gnedenko, B. W., und W. S. Koroljuk (Гнеденко, Б. В., и В. С. Королюк)

[1] О максимальном расхождении двух эмпирических распределений. Доклады Акад. Наук СССР **80** (1951), 525.

Gnedenko, B. W., und W. S. Michalewitsch (Гнеденко, Б. В., и В. С. Михалевич)

[1] О распределении числа выходов одной эмпирической функции распределения над другой. Доклады Акад. Наук СССР **82** (1952), 841.

Gnedenko, B. W., und E. L. Rwatschewa (Гнеденко, Б. В., и Е. Л. Рвачева)

[1] Об одной задаче сравнения двух эмпирических распределенний. Доклады Акад. Наук СССР **82** (1952), 513.

Good, I. J.

[1] The likelihood ratio test for Markov chains. Biometrika **42** (1955), 531; Corrigenda Biometrika **44** (1957), 301.

Goodman, L. A.

[1] Simplified run tests and likelihood ratio tests for Markov chains. Biometrika **45** (1958), 181.

[2] On some statistical test for mth order Markov chains. Annals of Math. Statistics **30** (1959), 154.

GOTKIN, L. G., und L. S. GOLDSTEIN
[1] Grundkurs in Statistik, Ein programmiertes Lehrbuch in zwei Bänden. Oldenbourg, München 1967 (Übersetzung aus dem Amerikanischen).
GRAF, U., und H.-J. HENNING
[1] Statistische Methoden bei textilen Untersuchungen. 3. Aufl., Springer-Verlag, Berlin—Göttingen—Heidelberg 1960.
GRANT, E. L.
[1] Statistical quality control. McGraw-Hill, New York 1956.
GRAYBILL, F. A.
[1] An introduction to linear statistical models. McGraw Hill Book Co., New York 1961.
GREENWOOD, M., and G. U. YULE
[1] An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. J. Royal Statist. Soc. **83** (1920), 255.
[2] The statistics of anti-typhoid and anti-cholera inoculations and the interpretation of such statistics in general. Proc. Royal Soc. Medicine **8** (1915), 113.
GRENANDER, U.
[1] On empirical spectral analysis of stochastic processes. Ark. Mat. **1** (1952), 503.
[2] Einführung in das Studium der mathematischen Statistik, VEB Fachbuchverlag, Leipzig 1969 (Übersetzung aus dem Schwedischen).
GRENANDER, U., and M. ROSENBLATT
[1] Statistical analysis of stationary time series. John Wiley, New York/Almqvist and Wiksell, Stockholm 1957.
GRUBBS, F. E., and C. L. WEAVER
[1] The best unbiased estimate of the population standard deviation based on group ranges. J. Amer. Stat. Assn. **42** (1947), 224.
GUILFORD, J. P.
[1] Fundamental statistics in psychology and education. McGraw-Hill, New York—Toronto—London 1956.
GUMBEL, E. J.
[1] The distribution of the range. Annals of Math. Statistics **18** (1947), 384.
[2] Statistics of extremes. Columbia University Press 1958.
[3] Multivariate distributions with given margins and analytical examples. Bull. Inst. Intern. Stat. **37**, Nr. 3 (1960), 3.
[4] Bivariate logistic distributions. J. Amer. Stat. Assoc. **56** (1961), 335.
GÜNTER, N. M., und R. O. KUSMIN
[1] Aufgabensammlung zur höheren Mathematik, Band II, 3. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1967 (Übersetzung aus dem Russischen).
HÁJEK, J.
[1] Prispevky k teorii statistichého odhadu, kandidátská disertace. Praha 1955.
[2] Linear estimation of the mean value of a stationary random process with convex correlation function. Czechoslovak. Math. J. **6** (1956), 94.
HÁJEK, J., and A. RÉNYI
[1] Generalization of an inequality of Kolmogorov. Acta Math. Acad. Sci. Hung. **6** (1955), 281.
HAJNAL, J.
[1] The ergodicity properties of non-homogeneous finite Markov chains. Proc. Cambridge Philos. Soc. **52** (1956), 67.

[2] Weak ergodicity in non-homogeneous Markov chains. Proc. Cambridge Philos. Soc. **54** (1958), 233.

HALD, A.

[1] Statistical tables and formulas. John Wiley & Sons, New York/Chapman & Hall, London 1952.

[2] Statistical theory with engineering applications. John Wiley & Sons, New York/Chapman & Hall, London 1952.

HALMOS, P. R.

[1] The theory of unbiased estimation. Annals of Math. Statistics **17** (1946), 34.

[2] Measure theory. Van Nostrand, Toronto—New York—London 1950.

HALMOS, P. R., and L. J. SAVAGE

[1] Application of the Radon-Nikodym theorem to the theory of sufficient statistics. Annals of Math. Statistics **20** (1949), 225.

HAMBURGER, H.

[1] Über eine Erweiterung des Stieltjesschen Momentenproblems. Math. Annalen **81** (1920), 235; **82** (1921), 120, 168.

HANNAN, E. T.

[1] Time series analysis. Methuen, London/John Wiley, New York 1960.

HARLEY, B. J.

[1] Relation between the distributions of the non-central t and of a transformed correlation coefficient. Biometrika **44** (1957), 219.

HARRIS, T. E.

[1] Branching processes. Annals of Math. Statistics **19** (1948), 474.

HARTLEY, H. O., and E. S. PAERSON

[1] The probability integral of the range in samples of n observations from a normal population. Biometrika **32** (1942), 301.

HASELOFF, O. W., und H.-J. HOFFMANN

[1] Kleines Lehrbuch der Statistik. 3. Aufl., de Gruyter, Berlin 1968.

HAUSDORFF, F.

[1] Beiträge zur Wahrscheinlichkeitsrechnung. Berichte Ver. Sächs. Gesell. Wiss. Leipzig, Math.-Phys. Klasse, **53** (1901), 152.

[2] Grundzüge der Mengenlehre. De Gruyter, Berlin 1937.

HEINHOLD, J., und K.-W. GAEDE

[1] Ingenieur-Statistik. Oldenbourg, München—Wien 1964.

HELMERT, F. R.

[1] Über die Wahrscheinlichkeit der Potenzsummen und über einige damit in Zusammenhang stehende Fragen. Z. Math. Phys. **21** (1876), 192.

HEPPES, A.

[1] On the determination of probability distribution of more dimensions by their projections. Acta Math. Acad. Sci. Hung. **7** (1956), 403.

HERGLOTZ, G.

[1] Über Potenzreihen mit positivem reellen Teil im Einheitskreis. Berichte Ver. Sächs. Gesell. Wiss. Leipzig, Math.-Phys. Klasse, **63** (1911), 501.

HIRSCHFELD, H. O.

[1] A connection between correlation and contingency. Proc. Cambridge Philos. Soc. **31** (1935), 520.

HODGES, J. L., jr.

[1] The significance probability of the Smirnov two-sample test. Ark. Mat. **3** (1957), 469.

HODGES, J. L., jr., and L. LECAM

[1] The Poisson approximation to the Poisson binomial distribution. Annals of Math. Statistics **31** (1960), 737.

HOEFFDING, W.

[1] "Optimum" non-parametric tests. Proc. Second Berkeley Symp. **1951**, p. 83.

HOEL, P. G.

[1] A test for Markov chains. Biometrika **41** (1954), 430.

HOGBEN, L.

[1] Zahl und Zufall. Oldenbourg, München 1956 (Übersetzung aus dem Englischen).

HORTON, H. B.

[1] A method for obtaining random numbers. Annals of Math. Statistics **19** (1948), 81.

HOSTINSKÝ, B.

[1] Méthodes générales du calcul des probabilités. Gauthier-Villars, Paris 1931.

HOTELLING, H.

[1] The consistency and ultimate distribution of optimum statistics. Trans. Amer. Math. Soc. **32** (1930), 360.

HSU, P. L.

[1] Contributions to the theory of Student's *t*-test. Stat. Res. Mem. **2** (1938), 1.

[2] A new proof of the joint product moment distribution. Proc. Cambridge Philos. Soc. **35** (1939), 336.

[3] Analysis of variance from the power function standpoint. Biometrika **32** (1941), 62.

HUNT, G. A.

[1] Markoff processes and potentials. Ill. J. Math. **1** (1957), 44; **2** (1958), 151.

[2] Markoff processes and Martin boundaries. Ill. J. Math. **4** (1960), 313.

HYMANS, S. H.

[1] Probability theory with applications to econometrics and decision-making. Prentice-Hall, Englewood Cliffs, N.J., 1967.

IBRAGIMOW, I. A. (Ибрагимов, И. А.)

[1] Некоторые предельные теоремы для стационарных в узком смысле вероятностных процессов. Доклады Акад. Наук СССР **125** (1959), 711.

IRWIN, J. O.

[1] Mathematical theorems involved in the analysis of variance. J. Royal Statist. Soc. **94** (1931), 284.

ISAACSON, S. L.

[1] On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters. Annals of Math. Statistics **22** (1951), 217.

ISHII, G.

[1] On the exact probabilities of Rényi's tests. Ann. Inst. Stat. Math. Tokyo **11** (1959), 17.

ITÔ, S.

[1] Brownian motion in a topological group and in its covering group. Rend. Circ. Mat. Palermo (2) **1** (1952), 40.

JAGLOM, A. M. (Яглом, А. М.)

[1] Введение в теорию стационарных случайных функции. Успехи Матем. Наук **7**: 5 (1952), 3.

[2] Einführung in die Theorie stationärer Zufallsfunktionen. Akademie-Verlag, Berlin 1959 (Übersetzung aus dem Russischen).

Jaglom, A. M., und I. M. Jaglom (Яглом, А. М., и. И. М. Яглом)

[1] Wahrscheinlichkeit und Information. 3. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1967 (Übersetzung aus dem Russischen).

Janko, J.

[1] Statistické tabulky. Čes. Slov. Akad. Věd., Praha 1958.

Jánossy, L., A. Rényi and J. Aczél

[1] On compound Poisson distributions. Acta Math. Acad. Sci. Hung. **1** (1950), 209.

Jeffreys, H.

[1] Theory of probability. The Clarendon Press, Oxford 1939.

Jenkins, G. M., and M. B. Priestley

[1] The spectral analysis of time series. J. Royal Statist. Soc., ser. B, **19** (1957), 1.

Jiřina, M.

[1] Stochastic branching processes with continuous state space. Czechoslovak. Math. J. **8** (1958), 292.

Johnson, N. L., and B. L. Welch

[1] Applications of the non-central t-distribution. Biometrika **31** (1939), 362.

Juschkewitsch, A. A. (Юшкевич, А. А.)

[1] О дифференцируемости переходных вероятностей однородного марковского процесса со счетным числом состояний. Уч. Зап. МГУ, вып. 186 (1959), 141.

Kac, M.

[1] On a characterization of the normal distribution. Amer. J. Math. **61** (1939), 726.

[2] On distributions of certain Wiener functionals. Trans. Amer. Math. Soc. **65** (1949), 1.

[3] On deviations between theoretical and empirical distributions. Proc. Nat. Acad. Sci. USA **35** (1949), 252.

[4] On some connections between probability theory and differential and integral equations. Proc. Second Berkeley Symp. 1951, p. 189.

[5] Probability and related topics in physical sciences. Interscience Publishers, London—New York 1959.

Kac, M., and J. F. Siegert

[1] An explicit representation of a stationary Gaussian process. Annals of Math. Statistics **18** (1947), 438.

Kadyroff, M. (Кадыров, М.)

[1] Таблицы случайных чисел. Ташкент 1936.

Kamke, E.

[1] Das Lebesgue-Stieltjes-Integral. 2. Aufl., B. G. Teubner, Leipzig 1960.

Karhunen, K.

[1] Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fennicae, ser. A, I, Math.-Phys., Nr. 37 (1947).

Karlin, S., and J. L. McGregor

[1] The differential equations of birth-and-death processes and the Stieltjes moment problem. Trans. Amer. Math. Soc. **85** (1957), 489.

[2] The classification of birth and death processes. Trans. Amer. Math. Soc. **86** (1957), 366.

[3] Linear growth birth and death processes. J. Math. Mech. **7** (1958), 643.

[4] Many server queueing processes with Poisson input and exponential service times. Pacific J. Math. **8** (1958), 87.

KAUCKY, J.
[1] Quelques remarques sur les chaînes de Markoff. Spisy Brnó 1930, Nr. 131.
KAWATA, T., and H. SAKAMOTO
[1] On the characterization of the normal population by the independence of the sample mean and the sample variance. J. Math. Soc. Japan **1** (1949), 111.
KEMENY, J. G., and J. L. SNELL
[1] Finite Markov chains. Van Nostrand, Princeton 1960.
[2] On Markov chain potentials. Annals of Math. Statistics **32** (1961), 709.
[3] Potentials for denumerable Markov chains. J. Math. Analysis and Appl. **3** (1961), 196.
KEMPTHORNE, O.
[1] The design and analysis of experiments. John Wiley, New York 1952.
KENDALL, D. G.
[1] Stochastic processes and population growth. J. Royal Statist. Soc., ser. B, **11** (1949), 230.
[2] On the generalized "birth and death" process. Annals of Math. Statistics **19** (1948), 1.
KENDALL, M. G.
[1] The advanced theory of statistics. Griffin, London 1948.
KENDALL, M. G., and B. SMITH
[1] Tables of random numbers. Tracts for computers Nr. 24, 1940.
KENDALL, M. G., and A. STUART
[1] The advanced theory of statistics I—III. Griffin, London 1958—1965.
KIEFER, J.
[1] K-sample analogues of the Kolmogorov-Smirnov and Cramér-von Mises tests. Annals of Math. Statistics **30** (1959), 420.
KIEFER, J., and L. WEISS
[1] Some properties of generalized sequential probability ratio tests. Annals of Math. Statistics **28** (1957), 57.
KIEFER, J., and J. WOLFOWITZ
[1] Sequential tests of hypotheses about the mean occurrence time of a continuous parameter Poisson process. Naval Research Logistics Quarterly **3**, Nr. 3 (1956), 205.
[2] Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Annals of Math. Statistics **27** (1956), 887.
[3] The deviation of the empiric distribution function of vector chance variables. Trans. Amer. Math. Soc. **87** (1958), 173.
KINNEY, J. R.
[1] Continuity properties of sample functions of Markov processes. Trans. Amer. Math. Soc. **74** (1953), 280.
KITAGAWA, T.
[1] Tables of Poisson distribution. Baifukan, Tokyo 1952.
KLEMM, L., H. J. RIEHL u. a.
[1] Statistische Kontrollmethoden in der Textilindustrie. 2. Aufl., Fachbuchverlag, Leipzig 1964.
KOLMOGOROFF, A. N. (Колмогоров, А. Н.; KOLMOGOROV, A. N.; KOLMOGOROW, A. N.)
[1] Über die Summen durch den Zufall bestimmter unabhängiger Größen. Math. Annalen **99** (1928), 309.
[2] Sur la loi forte des grands nombres. Comptes Rendus Paris **191** (1930), 910.
[3] Über die analytischen Methoden der Wahrscheinlichkeitsrechnung. Math. Annalen **104** (1931), 415.

[4] Sur le problème d'attente. Матем. Сб. **38** (1931), 101.
[5] Sulla forma generale di un processo stocastico omogeneo. Rend. Accad. Lincei, Cl. Sci. Fis. Mat. **15** (6) (1932), 805.
[6] Sulla determinazione empirica di una legge di distribuzione. Giorn. Ist. Ital. Attuari **4** (1933), 83.
[7] Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin 1933.
[8] La transformation de Laplace dans les espaces linéaires. Comptes Rendus Paris **200** (1935), 1717.
[9] Интерполирование и экстраполирование стационарных случайных последовательностей. Известия Акад. Наук, сер. матем., **5** (1941), 3.
[10] К вопросу о дифференцируемости переходных вероятностей в однородных по времени процессах Маркова со счетным числом состояний. Моск. Учен. зап. ун-та 148; Математика **4** (1951), 53.
[11] Цепи Маркова со счетным числом возможных состояний. Моск. Бюлл. ун-та (А) **1**: 3 (1937), 1.
[12] Кривые в гильбертовском пространстве, инвариантные по отношению к однопараметрической группе движений. Доклады Акад. Наук СССР **26** (1940), 6.
[13] Стационарные последовательности в гильбертовском пространстве. Моск. Бюлл. ун-та (А) **2**: 6 (1941), 1.
[14] Локальная предельная теорема для классических цепей Маркова. Известия Акад. Наук, сер. матем., **13** (1949), 287.
[15] Несмещенные оценки. Известия Акад. Наук, сер. матем., **14** (1950), 303.
[16] Das Gesetz des iterierten Logarithmus. Math. Annalen **101** (1929), 126.

KOLMOGOROFF, A. N., und N. A. DMITRIEW (Колмогоров, А. Н., и Н. А. Дмитриев)
[1] Ветвящиеся случайные процессы. Доклады Акад. Наук СССР **56** (1947), 7.

KOLMOGOROFF, A. N., und J. W. PROCHOROW (Колмогоров, А. Н., и Ю. В. Прохоров)
[1] Zufällige Funktionen und Grenzverteilungssätze, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 113.
[2] О суммах слачайного числа случайных слагаемых. Успехи Матем. Наук **4**: 4 (32) (1949), 168.

KOLODZIEJCZYK, S.
[1] On an important class of statistical hypotheses. Biometrika **27** (1935), 161.

KONEČNÝ, M.
[1] Sur la théorie des chaînes de Markoff. Spisy Brno 1931, Nr. 147.

KOOPMAN, B. O.
[1] On distributions admitting a sufficient statistic. Trans. Amer. Math. Soc. **39** (1936), 399.

KOROLJUK, W. S. (Королюк, В. С.)
[1] О расхождении эмпирических распределений для случая двух независимых выборок. Известия Акад. Наук, сер. матем., **19** (1953), 81.

KOTLARSKI, I.
[1] On random variables whose quotient follows the Cauchy law. Coll. Math. **7** (1960), 277.

KRETSCHMER, S. I. (Кречмер, С. И.)
[1] Исследование микропульсации температурного поля в атмосфере. Доклады Акад. Наук СССР **84** (1952), 55.

KREYSZIG, E.

[1] Statistische Methoden und ihre Anwendungen. 3. Aufl., Vandenhoeck & Ruprecht, Göttingen und Zürich 1968.

KRICKEBERG, K.

[1] Wahrscheinlichkeitstheorie. Teubner, Stuttgart 1963.

KRUSKAL, W. H.

[1] A nonparametric test for the several sample problem. Annals of Math. Statistics **23** (1952), 525.

KUBIK, L.

[1] O rozkładach granicznych sum r-punktowych zmiennych losowych. Prace Matematyczne **4** (1) (1960), 111.

KUIPER, N. H.

[1] Alternative proof of a theorem of Birnbaum and Pyke. Annals of Math. Statistics **30** (1959), 251.

[2] Tests concerning random points on a circle. Proc. Nederl. Akad. Wet. Amsterdam A **63**, Indagations Math. **22** (1960), 38.

KULLBACK, S.

[1] An application of characteristic functions to the distribution problem of statistics. Annals of Math. Statistics **5** (1934), 263.

KWIT, I. D. (Квит, И. Д.)

[1] О теореме Н. В. Смирнова относительно сравнения двух выборок. Доклады Акад. Наук СССР **71** (1950), 229.

LAHA, R. G.

[1] On a characterization of the Gamma distribution. Annals of Math. Statistics **25** (1954), 784.

[2] On a characterization of the multivariate normal distribution. Sankhya **14** (1955), 367.

[3] On some properties of the normal and gamma distributions. Proc. Amer. Math. Soc. **7** (1956), 172.

[4] On a characterization of the normal distribution from properties of suitable linear statistics. Annals of Math. Statistics **28** (1957), 126.

[5] An example of a non-normal distribution where the quotient follows the Cauchy law. Proc. Nat. Acad. Sci. USA **24** (1958), 222.

[6] On a class of distribution functions where the quotient follows the Cauchy law. Trans. Amer. Math. Soc. **93** (1959), 205.

LAMPERTI, J.

[1] An invariance principle in renewal theory. Technical Report Nr. 8, Appl. Math. Stat. Lab., Stanford University 1961.

LANGE, O.

[1] Statistical estimation of parameters in Markov processes. Coll. Math. **3** (1955), 147.

LAPLACE, P. S.

[1] Théorie analytique des probabilités. Paris 1820.

LECAM, L.

[1] On the asymptotic theory of estimation and testing hypotheses. Proc. Third Berkeley Symp. **1** (1956), 129.

[2] Convergence in distribution of stochastic processes. Univ. Calif. Publ. Stat. **2**, Nr. 11 (1957), 207.

[3] Les proprietes asymptotiques des solutions de Bayes. Publ. Inst. Stat. Univ. Paris **7**, Nr. 3—4 (1959), 17.

LEE, Y. W.

[1] Statistical theory of communication. New York—London 1960.

LEGENDRE, A. M.

[1] Nouvelles méthodes pour la détermination des orbites des comètes. Didot, Paris 1805.

LEHMANN, E. L.

[1] A general concept of unbiasedness. Annals of Math. Statistics **22** (1951), 587.

[2] Consistency and unbiasedness of certain non-parametric tests. Annals of Math. Statistics **22** (1951), 165.

[3] The power of rank tests. Annals of Math. Statistics **24** (1953), 23.

[4] Testing statistical hypotheses. John Wiley, New York—London 1959.

LEHMANN, E. L., and H. SCHEFFÉ

[1] Completeness, similar regions and unbiased estimation I. Sankhya **10** (1950), 305.

LEIMBACHER, W. R.

[1] On some classes of sequential procedures for obtaining confidence intervals of given length. Univ. Calif. Publ. Stat. **2**, Nr. 1 (1953), 1.

LEVENE, H.

[1] On the power function of tests of randomness based on runs up and down. Annals of Math. Statistics **23** (1952), 34.

LÉVY, P.

[1] Calcul des probabilités. Gauthier-Villars, Paris 1925.

[2] Sur les intégrals dont les éléments sont des variables aléatoires indépendantes. Annali R. Scu. Norm. Sup. Pisa **3** (2) (1934), 337.

[3] Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris 1937.

[4] Processus stochastiques et mouvement Brownien. Gauthier-Villars, Paris 1948.

[5] Systèmes Markoviens et stationnaires. Cas dénombrable. Ann. École Normale Sup. Paris Suppl. (3) **68** (1951), 327.

[6] Compléments à l'étude des processus de Markoff. Ann. École Normale Sup. Paris **69** (1952), 203.

[7] Wiener's random functions, and other Laplacian random functions. Proc. Second Berkeley Symp. 1951, p. 471.

[8] Sur certains processus stochastiques homogènes. Compositio Math. **7** (1939), 283.

[9] Le mouvement Brownien plan. Amer. J. Math. **62** (1940), 487.

[10] Systèmes Semi-Markoviens à au plus une infinité dénombrable d'états possibles. Proc. Int. Congr. Math. Amsterdam **2** (1954), 294; **3** (1954), 416.

[11] Le mouvement brownien à $n = 2p + 1$ paramètres I—III. Comptes Rendus Paris **239** (1954), 1181, 1584; **240** (1955), 1043.

LEXIS, W.

[1] Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik. Gustav Fischer, Jena 1903.

LIEBERMAN, G. J., and H. SOLOMON

[1] Multilevel continuous sampling plans. Annals of Math. Statistics **26** (1955), 686.

LINDEBERG, J. W.

[1] Eine neue Herleitung des Exponentialgesetzes der Wahrscheinlichkeitsrechnung. Math. Z. **15** (1922), 221.

LINDER, A.

[1] Statistische Methoden für Naturwissenschafter, Mediziner und Ingenieure. 4. Aufl., Birkhäuser, Basel 1964.

LINNIK, J. W. (Линник, Ю. В.)
[1] Линейные статистики и нормальный закон распределения. Доклады Акад. Наук СССР **83** (1952), 353.
[2] Одна задача о характеристических функциях вероятностных распределений. Успехи Матем. Наук **10**: 1 (63) (1955), 137.
[3] Die Methode der kleinsten Quadrate in moderner Darstellung. VEB Deutscher Verlag der Wissenschaften, Berlin 1961 (Übersetzung aus dem Russischen).

LJAPUNOFF, A. M.
[1] Sur une proposition de la théorie des probabilités. Bull. Acad. Sci. St. Pétersburg **13** (1900), 359.
[2] Nouvelle forme du théorème sur la limite de probabilité. Mem. Acad. Sci. St. Pétersburg **12**, Nr. 5 (1901).

LOÈVE, M.
[1] Sur les fonctions aléatoires stationnaires de second ordre. Rev. Sci. **83** (1945), 297.
[2] Fonctions aléatoires de second ordre. Comptes Rendus Paris **220** (1945), 380; **222** (1946), 469.
[3] On almost sure convergence. Proc. Second Berkeley Symp. 1951, p. 279.
[4] Probability theory. Van Nostrand, New York 1955 (3. ed., 1963).
[5] Ranking limit problems. Proc. Third Berkeley Symp. **2** (1956), 177.
[6] Études asymptotiques des sommes de variables aléatoires. J. de Math. **24** (1945), 249.

ŁOMNICKI, A.
[1] Nouveaux fondements du calcul des probabilités. Fundamenta Math. **4** (1923), 34.

ŁOMNICKI, Z. A., and S. K. ZAREMBA
[1] On estimating the spectral density function of a stochastic process. J. Royal Statist. Soc., ser. B, **19** (1957), 13.

LORCH, L., and D. J. NEWMAN
[1] The Lebesgue constants for regular Hausdorff methods. Canad. J. Math. **13** (1961), 283.

LUKACS, E.
[1] A characterization of the normal distribution. Annals of Math. Statistics **13** (1942), 91.
[2] Applications of Faà di Bruno's formula in mathematical statistics. Amer. Math. Monthly **62** (1955), 340.
[3] Characterization of populations by properties of suitable statistics. Proc. Third Berkeley Symp. **2** (1956), 215.
[4] Some extensions of a theorem of Marcinkiewicz. Pacific J. Math. 8 (1958), 487.
[5] Characteristic functions. Griffin, New York 1960.

ŁUKASZEWICZ, J., a W. SADOWSKI
[1] O porównywaniu kilku populacji z populacją kontrolną. Zastosowania Mat. **3** (1957), 204.

LUNDBERG, O.
[1] On random processes and their applications to sickness and accident statistics. University of Stockholm, Thesis, Uppsala 1940.

LUSTIG, H., J. PFANZAGL und L. SCHMETTERER
[1] Moderne Kontrolle. Wien 1956.

MADOW, W. G., and L. H. MADOW
[1] On the theory of systematic sampling. Annals of Math. Statistics **15** (1944), 1.

MAHALANOBIS, P. C.
[1] On large sample surveys. Phil. Trans. Royal Soc., ser. B, **231** (1944), 329.

Malécot, G.

[1] Sur un problème de probabilités en chaîne que pose la génétique. Comptes Rendus Paris **219** (1944), 379.

Malmquist, S.

[1] On a property of order statistics from a rectangular distribution. Skand. Aktuar. **33** (1950), 214.

[2] On certain confidence contours for distribution functions. Annals of Math. Statistics **25** (1954), 523.

Manija, G. M. (Мания, Г. М.)

[1] Обобщение критерия А. Н. Колмогорова для оценки закона распределения по эмпирическим данным. Доклады Акад. Наук СССР **69** (1949), 495.

Mann, H. B.

[1] Analysis and design of experiments. Dover, New York 1949.

Mann, H. B., and A. Wald

[1] On the choice of the number of class intervals in the application of the chi-square test. Annals of Math. Statistics **13** (1942), 306.

Mann, H. B., and D. R. Whitney

[1] On a test whether one of two random variables is stochastically larger than the other. Annals of Math. Statistics **18** (1947), 50.

Marbe, K.

[1] Die Gleichförmigkeit in der Welt. München 1916.

Marcinkiewicz, J.

[1] Sur une propriété de la loi de Gauss. Math. Z. **44** (1939), 612.

Marcinkiewicz, J., et A. Zygmund

[1] Quelques théorèmes sur les fonctions indépendantes. Studia Math. **7** (1938), 104.

Marczewski, E.

[1] Remarks on the Poisson stochastic process II. Studia Math. **13** (1953), 130.

Markoff, A. A. (Марков, А. А.)

[1] Распространение закона больших чисел на величины зависящие друг от друга. Известия Матем. Физ. Об. при Каз. Унив., сер. 2, **15** (1906), 135.

[2] Исследование замечательного случая зависимых испытаний. Известия Акад. Наук СПБ, сер. VI, **1** (1907), 61.

[3] Распространение предельных теорем исчисления вероятностей на сумму величин связанных в цепь. Mémoires de l'Académie St. Pétersburg, sér. 8, **22** (1908).

[4] Пример статистического исследования над текстом «Евгения Онегина» иллюстрирующий связь испытании в цепь. Известия Акад. Наук **7** (6) (1913).

[5] Исчисление вероятностей. Изд. 4, Москва 1924.

Masani, P.

[1] The prediction theory of multivariate stochastic processes III. Acad. Math. **140, Nr. 1—2** (1960), 141.

Masano, A.

[1] Optimization of stochastic systems. New York 1967.

Massey, F. J.

[1] A note on the power of a nonparametric test. Annals of Math. Statistics **21** (1950), 440.

[2] The distribution of the maximum deviation between two sample cumulative step functions. Annals of Math. Statistics **22** (1951), 125.

MAULDON, J. G.

[1] Characterizing properties of statistical distributions. Quarterly J. Math., ser. 2, **7** (1956), 155.

MAZURKIEWICZ, S.

[1] Podstawy rachunku prawdopodobieństwa. PWN, Warszawa 1956.

MCKEAN, H. P.

[1] Sample functions of stable processes. Annals Math., 2nd ser., **61** (1955), 564.

MEJZLER, D. G. (Мейзлер, Д. Г.)

[1] О предельном распределении максимального члена вариационого ряда. Доклады Акад. Наук УССР **1** (1950), 3.

MEJZLER, D. G., O. S. PARASIUK und E. L. RWATSCHEWA (Мейзлер, Д. Г., О. С. Парасюк и Е. Л. Рвачева)

[1] О многомерной локальной предельной теореме теории вероятностей. Успехи Матем. Наук **1** (1949), 9.

MERRINGTON, M., and E. S. PEARSON

[1] An approximation to the distribution of noncentral t. Biometrika **45** (1958), 484.

MICHALEWITSCH, W. S. (Михалевич, В. С.)

[1] О взаимном расположении двух эмпирических функций распределения. Доклады Акад. Наук СССР **85** (1952), 485.

MIHOC, G.

[1] Über verschiedene Ausdehnungen des Poissonschen Gesetzes auf endliche konstante Markoffsche Ketten, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 43.

VON MISES, R.

[1] Grundlagen der Wahrscheinlichkeitsrechnung. Math. Z. **4** (1919), 1.

[2] Wahrscheinlichkeit, Statistik und Wahrheit. Springer, Wien 1936.

[3] La distribution de la plus grande de n valeurs. Revue Math. de l'Union Interbalkanique **1**, Nr. 1 (1939), 141.

[4] On the correct use of Bayes' formula. Annals of Math. Statistics **13** (1942), 156.

[5] Über die Wahrscheinlichkeit seltener Ereignisse. Z. angew. Math. Mech. **1** (1921), 121.

[6] Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik. Franz Deuticke, Wien—Leipzig 1931.

MOOD, A. M.

[1] The distribution theory of runs. Annals of Math. Statistics **11** (1940), 367.

[2] Introduction to the theory of statistics. McGraw-Hill, New York 1950.

MORGENSTERN, D.

[1] Einführung in die Wahrscheinlichkeitsrechnung und mathematische Statistik. Springer-Verlag, Berlin—Göttingen—Heidelberg 1964.

MOSTELLER, F.

[1] A k-sample slippage test for an extreme population. Annals of Math. Statistics **19** (1948), 58.

MOURIER, E.

[1] Éléments aléatoires dans un espace de Banach. Thèse, Paris 1954.

MOYAL, J. E.

[1] Multiplicative population chains. Techn. Rep. No. 6, Appl. Math. Stat. Lab., Stanford University 1961.

[2] Multiplicative population processes. Techn. Rep. No. 7, Appl. Math. Stat. Lab., Stanford University 1961.

NAGAJEW, S. W. (Нагаев, С. В.)

[1] Некоторые предельные теоремы для однородных цепей Маркова. Теория вероятностей и ее применения **2** (1957), 389.

NASH, S. W.

[1] An extension of the Borel-Cantelli lemma. Annals of Math. Statistics **25** (1954), 165.

NATANSON, I. P. (Натансон, И. П.)

[1] Theorie der Funktionen einer reellen Veränderlichen. 2. Aufl., Akademie-Verlag, Berlin 1961 (Übersetzung aus dem Russischen).

NEYMAN, J.

[1] Zarys teorii i praktyki badania struktury ludności metoda reprezentacyjna. GUS, Warschau 1933.

[2] On the problem of confidence intervals. Annals of Math. Statistics **6** (1935), 111.

[3] Outline of a theory of statistical estimation based on the classical theory of probability. Phil. Trans. Royal Soc., ser. A, **236** (1937), 333.

[4] L'estimation statistique traitée comme un problème classique de probabilité. Actual. Scient. Ind. No. 739, Hermann, Paris 1938.

[5] Lectures and conferences on mathematical statistics. U. S. Department of Agriculture, Washington 1938.

[6] Basic ideas and some recent results of the theory of testing statistical hypotheses. J. Royal Statist. Soc. **105** (1944), 292.

[7] Su un teorema concernente le cosidette statistiche sufficienti. Giorn. Ist. Ital. Attuari **6** (1935), 320.

[8] Contributions to theory of the χ^2 test. Proc. Berkeley Symp. 1949, p. 239.

[9] First course in probability and statistics. New York 1950.

NEYMAN, J., and E. S. PEARSON

[1] On the use and interpretation of certain test criteria for purposes of statistical inference I, II. Biometrika **20** A (1928), 175, 263.

[2] On the problem of two samples. Bull. Int. Acad. Cracovie **A** (1930), 73.

[3] On the problem of the most efficient tests of statistical hypotheses. Phil. Trans. Royal Soc., ser. A, **231** (1933), 289.

[4] Contributions to the theory of testing statistical hypotheses. Stat. Res. Mem. **1** (1936), 1; **2** (1938), 25.

[5] Sufficient statistics and uniformly most powerful tests of statistical hypotheses. Stat. Res. Mem. **1** (1936), 113.

NEYMAN, J., and B. TOKARSKA

[1] Errors of the second kind in testing Student's hypothesis. J. Amer. Stat. Assn. **31** (1936), 318.

NICHOLSON, W. L.

[1] A computing formula for the power of the analysis of variance test. Annals of Math. Statistics **25** (1954), 607.

NOVÁK, J.

[1] Die topologische Struktur der Wahrscheinlichkeitsfelder, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 17.

OLKIN, I., and J. W. PRATT

[1] A multivariate Tchebysheff inequality. Annals of Math. Statistics **29** (1958), 226.

ONICESCU, O., and G. MIHOC

[1] Sur l'application de la notion de fonction caractéristique dans la théorie des chaînes de Markoff, Premier mémoire (cas discontinu). Matematica Cluj 1938.

[2] Sur les sommes de variables enchaînées. Second mémoire. Bull. Math. Soc. Roum. **41** (1939), 99.

OWEN, D. B.

[1] Handbook of statistical tables. Pergamon Press, London 1962.

OZOLS, V.

[1] Gnedenko-Koroluka teoremas vispãrinãjums uz tris izlasem pie divãm vienpuzigãn robezãm. Latvijas PRS Zinatum Akademijas Vestis **10** (1956), 141.

PALM, C.

[1] Intensitätsschwankungen im Fernsprechverkehr. Ericson Technics, Stockholm **44** (1943), 1.

PARZEN, E.

[1] On consistent estimates of the spectrum of a stationary time series. Annals of Math. Statistics **28** (1952), 329.

[2] Modern probability theory and its applications. John Wiley, New York—London 1960.

PAULSON, E.

[1] An optimum solution to the k-sample slippage problem for the normal distribution. Annals of Math. Statistics **23** (1952), 610.

PEARSON, E. S., and H. O. HARTLEY

[1] Biometrika tables for statisticians. Vol. I, 3rd ed., Vol. II, 1st ed., Cambridge University Press, Cambridge 1966, 1954.

PEARSON, K.

[1] Tables for the incomplete Γ-function. Cambridge University Press, London 1922.

[2] Tables for statisticians and biometricians I—II. Cambridge University Press, Cambridge 1924 bzw. 1931.

[3] Researches on the mode of distribution of the constants of samples taken at random from a bivariate normal population. Proc. Royal Soc. London A **112** (1926), 1.

[4] Tables of the incomplete B-function. Cambridge University Press, London 1934.

[5] On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil. Magazine V **50** (1900), 157.

[6] On the probable errors of the frequency constants. Biometrika **13** (1920—1921), 113.

[7] Contributions to the mathematical theory of evolution. Phil. Trans. Royal Soc., ser. A, **185** (1894), 71.

[8] Contributions to the mathematical theory of evolution, II. Skew variation in homogeneous material. Phil. Trans. Royal Soc., ser. A, **186** (1895), 343.

[9] On the systematic fitting of curves to observations and measurements I, II. Biometrika **1** (1901—1902), 265; **2** (1902—1903), 1.

PFANZAGL, J.

[1] Ein kombiniertes Test- und Klassifikations-Problem. Metrika **2**, Nr. 1 (1959), 11.

[2] Allgemeine Methodenlehre der Statistik I, II. 4. bzw. 3. Aufl., de Gruyter, Berlin 1967 bzw. 1968.

PITMAN, E. J. G.
[1] The closest estimation of statistical parameters. Proc. Cambridge Philos. Soc. **33** (1937), 212.
[2] On the derivatives of a characteristic function at the origin. Annals of Math. Statistics **27** (1956), 1156.

POINCARÉ, H.
[1] Calcul des probabilités. Gauthier-Villars, Paris 1912.

POISSON, S. D.
[1] Recherches sur la probabilité des jugements en matière criminelle et en matière civile. Paris 1837.

PÓLYA, G.
[1] Sur quelques points de la théorie des probabilités. Ann. Inst. H. Poincaré **1** (1930), 117.
[2] Herleitung des Gaußschen Satzes aus einer Funktionalgleichung. Math. Z. **18** (1923), 96.
[3] Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. Math. Z. **8** (1920), 171.

PRÉKOPA, A.
[1] On composed Poisson distributions IV. Acta Math. Acad. Sci. Hung. **3** (1952), 317.

PROCHOROW, J. W. (Прохоров, Ю. В.)
[1] Об усиленном законе больших чисел. Доклады Акад. Наук СССР **69** (1949), 607.
[2] Распределения вероятностей в функциональных пространствах. Успехи Матем. Наук **8**: 3 (1953), 165.
[3] Сходимость случайных процессов и предельные теоремы теории вероятностей. Теория вероятностей и ее применения **1** (1956), 177.
[4] Усиленная устойчивость сумм и неограниченно делимые распределения. Теория вероятностей и ее применения **3** (1958), 513.
[5] Несколько замечаний к усиленному закону больших чисел. Теория вероятностей и ее применения **4** (1959), 215.

PROCHOROW, J. W., und M. FISZ (Прохоров, Ю. В.)
[1] Характеристическое свойство нормального распределения в гильбертовом пространстве. Теория вероятностей и ее применения **2** (1957), 475.

PRZYBOROWSKI, J., and H. WILEŃSKI
[1] Homogeneity of results in testing samples from Poisson series. Biometrika **31** (1939 bis 1940), 313.

PYKE, R.
[1] The supremum and infimum of the Poisson process. Annals of Math. Statistics **30** (1959), 568.
[2] Markov renewal processes: definitions and preliminary properties. Annals of Math. Statistics **32** (1961), 1231.
[3] Markov renewal processes with finitely many states. Annals of Math. Statistics **32** (1961), 1243.

RAIKOW, D. A. (Райков, Д. А.)
[1] О разложении законов Гаусса и Пуассона. Известия Акад. Наук, сер. матем., **2** (1938), 91.

RAJCHMAN, A.
[1] Zaostrzone prawo wielkich liczb. Mathesis Polska **6** (1932), 145.

Ranga-Rao, R.
[1] Relations between weak and uniform convergence of measures with applications. Annals of Math. Statistics **33** (1962), 659.
Ranga-Rao, R., and V. S. Varadarajan
[1] On the decomposition of Haar measure in compact groups. Fundamenta Math. **49** (1961), 119.
Rao, C. R.
[1] Information and accuracy obtainable in an estimation of a statistical parameter. Bull. Calcutta Math. Soc. **37** (1945), 81.
[2] Linear statistical inference and its applications. Wiley & Sons, New York—London—Sidney 1968.
Rasch, D.
[1] Elementare Einführung in die mathematische Statistik. 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1970.
Reiersöl, O.
[1] Confluence analysis by means of instrumental sets of variables. Ark. Math. **32** A, Nr. 4 (1945).
Rényi, A.
[1] On some problems concerning Poisson processes. Publicationes Math. **2** (1957), 66.
[2] On composed Poisson distributions II. Acta Math. Acad. Sci. Hung. **2** (1951), 83.
[3] On the theory of order statistics. Acta Math. Acad. Sci. Hung. **4** (1953), 191.
[4] On a new axiomatic theory of probability. Acta Math. Acad. Sci. Hung. **6** (1955), 285.
[5] On projections of probability distributions. Acta Math. Acad. Sci. Hung. **3** (1952), 131.
[6] A characterization of Poisson process. Publ. Math. Inst. Hung. Acad. Sci. **1** (1956), 526.
[7] On the asymptotic distribution of the sum of a random number of independent random variables. Acta Math. Acad. Sci. Hung. **8** (1957), 193.
[8] New version of the probabilistic generalization of the large sieve. Acta Math. Acad. Sci. Hung. **10** (1959), 217.
[9] On measures of dependence. Acta Math. Acad. Sci. Hung. **10** (1959), 441.
[10] Neues Kriterium zum Vergleich zweier Stichproben. Magyar Tud. Akad. Mat. Inst. **2** (1953), 243.
[11] Wahrscheinlichkeitsrechnung, mit einem Anhang über Informationstheorie. 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1966.
[12] Briefe über die Wahrscheinlichkeit. VEB Deutscher Verlag der Wissenschaften, Berlin 1969 (Übersetzung aus dem Ungarischen).
Resnikoff, G. J., and G. J. Lieberman
[1] Tables of the noncentral t-distribution. Appl. Math. Stat. Lab., Stanford University, California 1957.
Reuter, G. E. H., and W. Ledermann
[1] On the differential equations for the transition probabilities of Markov processes with enumerably many states. Proc. Cambridge Philos. Soc. **49** (1953), 247.
Richter, H.
[1] Wahrscheinlichkeitstheorie. 2. Aufl., Springer-Verlag, Berlin—Heidelberg—New York 1966.
Richter, W.
[1] Локальные предельные теоремы для больших уклонений. Теория вероятностей и ее применения **2** (1957), 214.

[2] Многомерные локальные предельные теоремы для больших уклонений. Теория вероятностей и ее применения **3** (1958), 107.

RIESZ, F., und B. SZ.-NAGY

[1] Vorlesungen über Funktionalanalysis. 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1968 (Übersetzung aus dem Französischen).

RIJKOORT, P. J.

[1] A generalization of Wilcoxon's test. Proc. Royal Inst. Acad. Sci., ser. A, **55** (1952), 394.

RIORDAN, J.

[1] Stochastic service systems. John Wiley, New York 1962.

ROBBINS, H.

[1] The asymptotic distribution of the sum of a random number of random variables. Bull. Amer. Math. Soc. **54** (1948), 1151.

[2] Convergence of distributions. Annals of Math. Statistics **19** (1948), 72.

[3] On the equidistribution of sums of independent random variables. Proc. Amer. Math. Soc. **4** (1953), 786.

[4] An empirical Bayes' approach to statistics. Proc. Third Berkeley Symp. **1** (1956), 157.

ROMANOWSKY, W. I. (ROMANOVSKI, W.; Романовский, В. И.)

[1] On the moments of standard deviations and correlation coefficient in samples from normal populations. Metron **5**, Nr. 4 (1925), 3.

[2] On the distribution of the regression coefficient in samples from normal populations. Известия Акад. Наук **20** (1929), 643.

[3] Применение математической статистики в опытном деле. Москва—Ленинград 1947.

[4] Дискретные цепи Маркова. Москва—Ленинград 1949.

ROSANOW, J. A. (Розанов, Ю. А.)

[1] Спектральная теория многомерных стационарных случайных процессов с дискретным временем. Успехи Матем. Наук **13**: 2 (1958), 93.

[2] Wahrscheinlichkeitstheorie. Akademie-Verlag, Berlin/Pergamon Press, Oxford/Vieweg + Sohn, Braunschweig 1970 (Übersetzung aus dem Russischen).

ROSENBAUM, S.

[1] Tables for a nonparametric test of dispersion. Annals of Math. Statistics **24** (1953), 663.

ROSENBLATT, M.

[1] Some regression problems in time series. Proc. Third Berkeley Symp. **1** (1956), 165.

[2] A multidimensional prediction problem. Ark. Math. **3** (1958), 407.

RUBIN, H., and H. G. TUCKER

[1] Estimating the parameters of a differential process. Annals of Math. Statistics **30** (1959), 641.

RWATSCHEWA, E. L. (Рвачева, Е. Л.)

[1] О максимальном расхождении между двумя эмпирическими распределениями Укр. Матем. Ж. **4** (1952), 373.

RYLL-NARDZEWSKI, C.

[1] On the non-homogeneous Poisson process I. Studia Math. **14** (1953), 124.

[2] Remarks on the Poisson stochastic process III. Studia Math. **14** (1953), 314.

SADOWSKI, W., et al.

[1] Tablice statystyczne. PWN, Warszawa 1957.

[2] O nieparametrycznym tescie na porównywanie rozsiewów. Zastosowania Mat. **2** (1955), 161.

SAKS, S., and A. ZYGMUND

[1] Analytic functions. Polish Math. Soc., Warzawa 1952.

Sarymsakow, T. A. (Сарымсаков, Т. А.)

[1] Об эргодическом принципе для неоднородных цепей Маркова. Доклады Акад. Наук СССР **90** (1953), 25.

Savage, J. R.

[1] Contribution to the theory of Rank Order Statistics — the two-sample case. Annals of Math. Statistics **27** (1956), 590.

Savage, L. J.

[1] The foundations of statistics reconsidered. Proc. Fourth Berkeley Symp. **1** (1961), 575.

Schaafsma, A. H., und F. G. Willemze

[1] Moderne Qualitätskontrolle. 2. Aufl., Philips' Techn. Bibliothek, Eindhoven 1959.

Scheffé, H.

[1] The analysis of variance. John Wiley, New York 1959.

Schmetterer, L.

[1] Einführung in die mathematische Statistik. 2. Aufl., Springer-Verlag, Wien 1966.

[2] Bemerkungen zur Theorie der erwartungstreuen Schätzfunktionen. Mitteilungsblatt für Math. Statistik **9**, Nr. 2 (1957), 147.

[3] Über nichtparametrische Methoden in der mathematischen Statistik. Jahresber. Deutsch. Math.-Ver. **61**, Nr. 3 (1959), 104.

Schmid, P.

[1] On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions. Annals of Math. Statistics **29** (1958), 1011.

Seth, G. R.

[1] On the variance of estimates. Annals of Math. Statistics **20** (1949), 1.

Sewastjanow, B. A. (Севастьянов, Б. А.)

[1] Теория ветвящихся случайных процессов. Успехи Матем. Наук **6**: 6 (46) (1951), 47.

Siegel, S.

[1] Nonparametric statistics. New York 1956.

Sirashdinow, S. Ch. (Сираждинов, С. Х.)

[1] Уточнение предельных теорем для однородных цепей Маркова. Доклады Акад. Наук СССР **84** (1952), 1143.

Sittig, J., en H. Freudenthal

[1] De juiste maat, Leiden 1951.

Skitowitsch, W. P. (Скитович, В. П.)

[1] Линейные формы от независимых случайных величин и нормальный закон распределения. Известия Акад. Наук, сер. матем., **18** (1954), 185.

Skorochod, A. W. (Скороход, А. В.)

[1] О предельном переходе от последовательности сумм независимых случайных величин к однородному случайному процессу с независимыми приращениями. Доклады Акад. Наук СССР **104** (1955), 364.

[2] Предельные теоремы для процессов Маркова. Теория вероятностей и ее применения **3** (1958), 217.

Slutski, E. E. (Slutsky, E. E.; Слуцкий, Е. Е.)

[1] Über stochastische Asymptoten und Grenzwerte. Metron **5**, Nr. 3 (1925), 3.

[2] Alcuni proposizioni sulla teoria degli funzioni aleatorie. Giorn. Ist. Ital. Attuari **8** (1937), 183.

[3] Sur les fonctions aléatoires prèsque périodique et sur la décomposition des fonctions aléatoires stationnaires en composantes. Actual. sci. industr. 738, Herman, Paris 1938, p. 33.

Smirnow, N. W. (Smirnoff, N. W.; Смирнов, Н. В.)

[1] Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках. Моск. Бюлл. ун-та (A) **2**, Nr. 2 (1939), 3.

[2] Приближение законов распределения случайных величин по эмпирическим данным. Успехи Матем. Наук **10** (1944), 179.

[3] Предельные законы распределения для членов вариационного ряда. Труды Матем. ин-та Акад. Наук **25** (1949), 1.

Smirnow, N. W., und I. W. Dunin-Barkowski (Смирнов, Н. В., и. И. В. Дунин-Барковский)

[1] Mathematische Statistik in der Technik. 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1969 (Übersetzung aus dem Russischen).

Smirnow, W. I. (Смирнов, В. И.)

[1] Lehrgang der höheren Mathematik, Teil III/1. 5. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1967 (Übersetzung aus dem Russischen).

[2] Lehrgang der höheren Mathematik, Teil V. 2. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1967 (Übersetzung aus dem Russischen).

Smith, W. L.

[1] Regenerative stochastic processes. Proc. Royal Soc. London A **232** (1955), 6.

Smoluchowski, M.

[1] Über Brownsche Molekularbewegung. Ann. Phys. 48 (1915), 1103.

Snell, J. L.

[1] Application of martingale system theorems. Trans. Amer. Math. Soc. **73** (1952), 293.

Spitzer, F.

[1] A combinatorial lemma and its application to probability theory. Trans. Amer. Math. Soc. **82** (1956), 323.

Srivastava, A. B. L.

[1] Effect of non-normality on the power function of t-test. Biometrika **45** (1958), 421.

Stein, C.

[1] A two-sample test for a linear hypothesis whose power is independent of the variance. Annals of Math. Statistics **16** (1945), 243.

[2] A note on cumulative sums. Annals of Math. Statistics **17** (1946), 498.

[3] Unbiased estimates with minimum variance. Annals of Math. Statistics **21** (1950), 406.

Steinhaus, H.

[1] Les probabilités dénombrables et leur rapport à la théorie de la mesure. Fundamenta Math. **4** (1923), 286.

[2] Quality control by sampling (A plea for Bayes rule). Coll. Math. **2** (1950), 98.

[3] Tablice liczb przetasowanych. Math. Inst. Pol. Acad. Sci., Warszawa 1954.

[4] Über einige prinzipielle Fragen der mathematischen Statistik, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 55.

Stepanow, W. E. (Степанов, В. Е.)

[1] Некоторые статистические критерии для цепей Маркова. Теория вероятностей и ее применения **2** (1957), 143.

STEVENS, W. L.
[1] Distribution of groups in a sequence of alternatives. Ann. Eugenics Cambridge **9** (1939), 10.
STIELTJES, J.
[1] Recherches sur les fractions continues. Oeuvres complètes, Groningen. Amsterdam 1918.
STORM, R.
[1] Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle. 3. Aufl., VEB Fachbuchverlag, Leipzig 1969.
STUDENT
[1] The probable error of a mean. Biometrika **6** (1908), 1.
SUKHATME, P. V.
[1] Contribution to the theory of the representative method. J. Royal Statist. Soc., Suppl. **2** (1935), 253.
[2] A contribution to the problem of two samples. Proc. Indian Acad. Sci., A **2** (1936), 584.
[3] Sampling theory of surveys with applications. State College Press, Ames, Iowa, 1954.
SVERDRUP, E.
[1] The limit distribution of a continuous function of random variables. Memorandum fra Universitetets Socialøkonomiske Inst., Oslo 1951.
SWED, F. S., and C. EISENHART
[1] Tables for testing randomness of grouping in a sequence of alternatives. Annals of Math. Statistics **14** (1943), 66.
SYSKI, R.
[1] Introduction to congestion theory in telephone systems. Oliver and Boyd, Edinburgh 1960.
TAKÁCS, L.
[1] Occurrence and coincidence phenomena in case of happenings with arbitrary distribution law of duration. Acta Math. Acad. Sci. Hung. **2** (1951), 275.
[2] On secondary processes generated by a Poisson process and their applications in physics. Acta Math. Sci. Hung. **5** (1954), 203.
[3] On processes of happenings generated by means of a Poisson process. Acta Math. Acad. Sci. Hung. **6** (1955), 81.
[4] On stochastic processes connected with certain physical recording apparatures. Acta Math. Acad. Sci. Hung. **6** (1955), 363.
[5] On a sojourn time problem. Теория вероятностей и ее применения **3** (1958), 61.
[6] Introduction to the theory of queues. Oxford University Press, Oxford 1962.
[7] Stochastische Prozesse. Aufgaben und Lösungen. R. Oldenbourg Verlag, München—Wien 1966 (Übersetzung aus dem Englischen).
TANG, P. C.
[1] The power function of the analysis of variance tests with tables and illustrations of their use. Stat. Res. Mem. **2** (1938), 126.
TEICHER, H.
[1] On the convergence of projected distributions. Ann. Inst. Stat., Math. **9** (1958), 79.
THOMPSON, W. R.
[1] On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. Annals of Math. Statistics **7** (1936), 122.
TINTNER, G.
[1] Multiple regression for systems of equations. Econometrica **14** (1946), 5.

TIPPETT, L. H. C.

[1] On the extreme individuals and the range of samples taken from a normal population. Biometrika **17** (1925), 364.

[2] Random sampling numbers. Tracts for computers, Nr. 15, 1927.

TOCHER, K. D.

[1] Extension of the Neyman-Pearson theory of tests to discontinuous variates. Biometrika **37** (1950), 130.

TSAO, C. K.

[1] An extension of Massey's distribution of the maximum deviation between two-sample cumulative step functions. Annals of Math. Statistics **25** (1954), 587.

TSCHEBYSCHEFF, P. L. (Чебышев, П. Л.)

[1] О средних величинах. Полное собрание сочинений, т. 2. Москва—Ленинград 1948.

TSCHUPROV, A. A.

[1] On the mathematical expectation of the moments of the frequency distributions in case of correlated observations. Metron **2**, Nr. 4 (1923), 646.

TUCKER, H. G.

[1] A generalization of the Glivenko-Cantelli theorem. Annals of Math. Statistics **30** (1959), 1267.

TUMANJAN, S. CH. (Туманян, С. Х.)

[1] Асимптотическое распределение критерия χ^2 при одновременном возрастании объема наблюдений и числа групп. Теория вероятностей и ее применения **1** (1956), 131.

UHLMANN, W.

[1] Zu einem nichtparametrischen Test von E. L. Lehmann. Metrika **2**, Nr. 3 (1959), 169.

URBANIK, K.

[1] Limit properties of homogeneous Markoff processes with a denumerable set of states. Bull. Pol. Acad. Sci. **2** (1954), 371.

[2] Об одной задаче из теории процессов размножения и гибели. Acta Math. Acad. Sci. Hung. **7** (1956), 99.

USPENSKY, J. V.

[1] Introduction to mathematical probability. McGraw-Hill, New York 1937.

VAN DER VAART, H. R.

[1] Some remarks on the power function of Wilcoxon's test for the problem of two samples I, II. Proc. Royal Inst. Acad. Sci., ser. A, **53** (1950), 494, 507.

[2] On the characteristic functions of absolutely continuous distribution functions. Annals of Math. Statistics **33** (1962), 824.

VARADARAJAN, V. S.

[1] Weak convergence of measures on separable metric spaces. Sankhya **19** (1958), 15.

[2] On the convergence of sample probability distributions. Sankhya **19** (1958), 23.

VENN, J.

[1] The logic of chance. Recent edition, Chelsea, New York 1962.

VESSEREAU, A.

[1] Sur les conditions d'application de criterion χ^2 de Pearson. Rev. Stat. Appl. **6**, Nr. 2 (1958), 83.

VIELROSE, E.

[1] Tablice liczb losowych. GUS, Warszawa 1951.

VINCZE, I.
[1] Einige zweidimensionale Verteilungs- und Grenzverteilungsgesetze in der Theorie der geordneten Stichproben. Publ. Math. Inst. Hung. Acad. Sci. **2** (1957), 183.
[2] On some joint distributions and joint limiting distributions in the theory of order statistics. Publ. Math. Inst. Hung. Acad. Sci. **4** (1959), 29.

VOGEL, W.
[1] Eine allgemeine Klasse von Zwei-Personen-Spielen, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 73.

VAN DER WAERDEN, B. L.
[1] Testing a distribution function. Nederl. Akad. Wetensch. Proc., Ser. A, **56** = Indagationes Math. **15** (1953), 201.
[2] Order tests for the two-sample problem I—III. Nederl. Akad. Wetensch. Proc., Ser. A, **55**, **56** = Indagationes Math. **14** (1952), 453; **15** (1953), 303, 311.
[3] Mathematical statistics, Springer, Berlin—Heidelberg—New York 1969.

WALD, A.
[1] On cumulative sums of random variables. Annals of Math. Statistics **15** (1944), 283.
[2] Differentiation under the expectation sign in the fundamental identity of sequential analysis. Annals of Math. Statistics **17** (1946), 493.
[3] Sequential analysis. John Wiley, New York/Chapman and Hall, London 1947.
[4] Statistical decision functions. John Wiley, New York 1950.
[5] An extension of Wilk's method for setting tolerance limits. Annals of Math. Statistics **14** (1943), 45.
[6] Note on the consistency of the maximum likelihood estimate. Annals of Math. Statistics **20** (1949), 595.

WALD, A., and J. WOLFOWITZ
[1] On a test whether two samples are from the same population. Annals of Math. Statistics **11** (1940), 147.
[2] Confidence limits for continuous distribution functions. Annals of Math. Statistics **10** (1939), 105.
[3] Tolerance limits for a normal distribution. Annals of Math. Statistics **17** (1946), 208.
[4] Optimum character of the sequential probability ratio test. Annals of Math. Statistics **19** (1948), 326.

WALLIS, W. A.
[1] Techniques of statistical analysis. Statistical Research Group, Columbia University, New York 1946, chap. 17.
[2] Tolerance intervals for linear regression. Proc. Second Berkeley Symp. 1951, p. 43.

WALLIS, W. A., and H. V. ROBERTS
[1] Statistics: A new approach. Methuen, London 1957.

WALSH, J.
[1] Handbook of nonparametric statistics I, II. D. van Nostrand, London 1965, 1966.

WANG, S. J.
[1] On the limiting distribution of the ratio of two empirical distributions. Acta Math. Sinica **5** (1955), 253.

WATSON, G. S.
[1] The χ^2 goodness-of-fit test for normal distributions. Biometrika **44** (1957), 336.

[2] On chi-square goodness-of-fit tests for continuous distributions. J. Royal Statist. Soc. B **20** (1958), 44–61; Discuss. 61–72.

WEBER, E.

[1] Das Rückschlußproblem in der mathematischen Statistik, in: Bericht Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 81.

[2] Grundriß der biologischen Statistik, 6. Aufl., VEB Gustav Fischer Verlag, Jena 1967.

WELCH, B. L.

[1] The significance of the difference between two means when the population variances are unequal. Biometrika **29** (1938), 350.

WIENER, N.

[1] Differential space. J. Math. Phys. Math. Inst. Techn. **2** (1923), 131.

[2] Extrapolation, interpolation and smoothing of stationary time series, with engineering applications. Cambridge–New York 1949.

[3] Generalized harmonic analysis. Acta Math. **55** (1930), 117.

WIENER, N., and P. MASANI

[1] The prediction theory of multivariate stochastic processes I–II. Acta Math. **98** (1957), 111; **99** (1958), 93.

WILCOXON, F.

[1] Individual comparisons by ranking methods. Biometrics Bull. **1** (1945), 80.

WILKS, S. S.

[1] Determination of sample size for setting tolerance limits. Annals of Math. Statistics **12** (1941), **91**.

[2] Certain generalizations in the analysis of variance. Biometrika **24** (1932), 471.

[3] Mathematical statistics. John Wiley & Sons, New York–London 1962.

WILLIAMS, C. A.

[1] On the choice of the number and width of classes for the chi-square test of goodness of fit. J. Amer. Stat. Assn. **45** (1950), 77.

WISHART, J.

[1] The generalized product-moment distribution in samples from a normal multivariate population. Biometrika **20** A (1928), 32.

WISHART, J., and H. O. HIRSCHFELD

[1] A theorem concerning the distribution of joins between line segments. J. London Math. Soc. **11** (1936), 227.

WISZNIEWSKI, W.

[1] Pogoda naszych zim. Gazeta Observatora PIHM **1**, Nr. 3 (1948).

WOLD, H.

[1] A study in the analysis of stationary time series. Almquist and Wiksell, Uppsala 1938.

WOLFOWITZ, J.

[1] Additive partition functions and a class of statistical hypotheses. Annals of Math. Statistics **13** (1942), 247.

[2] The efficiency of sequential estimates and Wald's equation for sequential processes. Annals of Math. Statistics **18** (1947), 215.

[3] On Wald's proof of the consistency of the maximum likelihood estimate. Annals of Math. Statistics **20** (1949), 602.

[4] Nonparametric statistical inference. Proc. Berkeley Symp. 1949, p. 93.

[5] Generalization of a theorem of Glivenko-Cantelli. Annals of Math. Statistics **25** (1954), 131.

[6] Estimation by the minimum distance method in nonparametric stochastic difference equations. Annals of Math. Statistics **25** (1954), 203.
[7] The minimum distance method. Annals of Math. Statistics **28** (1957), 75.
[8] Convergence of the empiric distribution function on half spaces. Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling. Stanford, University Press 1960, p. 198.

WOODS, W. N., and A. T. BHARUCHA-REID
[1] Age dependent branching stochastic processes in cascade theory II. Il Nuovo Cimento, ser. 10, **10** (1958), 569.

YULE, G. U.
[1] A mathematical theory of evolution, based on conclusions of Dr. J. C. Willis, FRS. Phil. Trans. Royal Soc. London, ser. B, **213** (1924), 21.
[2] On the association of attributes in statistics. Phil. Trans. A **194** (1900), 257.
[3] On the methods of measuring the association between two attributes. J. Royal Statist. Soc. **75** (1912), 579.

ZINGER, A. A. (Зингер, А. А.)
[1] О независимых выборках из нормальной совокупности. Успехи Матем. Наук **6**: 5 (45) (1951), 172.

ZINGER, A. A., und J. W. LINNIK (Зингер, А. А., и Ю. В. Линник)
[1] Об одном аналитическом обобщении теоримы Крамера и его применения. Ленингр. Вестник ун-та **1** (1955), 51.

ZUBRZYCKI, S.
[1] Remarks on random, stratified and systematic sampling in a plane. Coll. Math. **6** (1958), 251.

ZYGMUND, A.
[1] A remark on characteristic functions. Annals of Math. Statistics **18** (1947), 272.
[2] A remark on characteristic functions. Proc. Second Berkeley Symp. 1951, p. 369.

Biometrisches Wörterbuch, VEB Deutscher Landwirtschaftsverlag, Berlin 1968.
Mathematik für die Praxis, Ein Handbuch. Band III, 3. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1966.
Wahrscheinlichkeitsrechnung und mathematische Statistik, Lexikon. Akademie-Verlag, Berlin 1970.

NAMENREGISTER

SACHREGISTER

